The Human Efficiency Evaluator
A tool to predict and explore monitoring behaviour

Sebastian Feuerstack*, Bertram Wortelen**

** Human Centered Design, OFFIS e.V. – Institute for Information Technology Oldenburg, Germany (e-mail: feuerstack@offis.de)*
*** Cognitive Psychology, C.v.O. University Oldenburg Oldenburg, Germany (e-mail: bertram.wortelen1@uni-oldenburg.de)*

Abstract: With more and more systems and machines operating autonomously, the role of the operator is changing from "being actively in control" to "monitor and intervene". Human Machine Interfaces (HMI) therefore need to be optimized so that they can efficiently be monitored by a human. We propose the Human Efficiency Evaluator (HEE), a software tool for (1) evaluating the impact of HMI design changes on the visual monitoring behavior of the operator and (2) to explore differences in understandings between a group of collaborating HMI designers or between HMI designers and their targeted audience: the operators. We describe the tool and highlight the model exploration capabilities of the HEE by reporting about two use cases: one in the maritime domain, in which the tool supported an HMI designer to get insights into human operators' monitoring behavior, and one in the automotive domain, in which the tool was used to reveal differences in understanding between six Human Factor Experts about the impact of three HMI design variants of an Urban ACC.

Keywords: Visual Attention, Human Factors, Supervisory Control, Software-supported method, Cognitive Modeling, Safety.

1. INTRODUCTION

The way we interact with machines is changing: Machines are getting continuously smarter and more and more are able to run autonomously without any human control. While still a human remains responsible for what the machine is doing, the human's role changes from "being in control" to "monitor and intervene". Human machine interfaces (HMI) therefore need to be optimized to support efficient human monitoring. Methods for analyzing HMIs for monitoring based on cognitive methods have already been used in safety-critical system design, e.g. for airplane cockpits [Wickens et al. (2003)], air traffic control [Wickens et al. (2008)] or in a clinical operation theater in a hospital [Koh et al. (2011)] by experts in human factors (HF) and cognitive engineering. Although these methods have been proven to be helpful, they have only been applied to a small extend so far. Reasons are the complexity of cognitive modelling, and a relatively high cost-benefit ratio.

We propose the Human Efficiency Evaluator (HEE), which is a software tool that eases the data collection, prediction, and exploration of monitoring behaviour. It enables human factor (HF) experts as well as HMI designers or domain experts with no or little background in cognitive modelling to model, predict and simulate human monitoring behaviour. Earlier versions of this tool were also used to predict cognitive and motor workload [Feuerstack et al. (2015)]. The tool can also combine and aggregate multiple monitoring models that were specified by an arbitrary large number of HF experts or designers. Prior work had indicated that although individual models differ, the model quality improves by averaging these models [Feuerstack & Wortelen (2016)]. The observed modelling variance in earlier studies gave us the idea that the tool can also be used to discover these differences in understanding of the working domain and the proposed HMI design within a group of peopled involved in the HMI design or between the HMI designers and the targeted audience (i.e. the operators with domain expertise). In this contribution we present the model exploration capabilities of the Human Efficiency Evaluator and report about two use cases, in which the tool has been applied: one in the automotive domain and one in the maritime domain.

2. VISUAL ATTENTION MODELING

The vision of human modeling is to provide methods, techniques and tools to generate predictions of human performance. The SEEV model of attention allocation [Wickens et al. (2001)] provides such a promising theory for modeling visual attention. It describes that "the allocation of attention in dynamic environments is driven by bottom up attention capture of *salient* events, which are inhibited by the *effort* required to move attention, and also driven by the *expectancy* of seeing *valuable* events" [McCarley et al. (2002)]. The SEEV model is used to predict the percentage of time, that someone spends looking at an area of interest (AOI). It is typically applied by HF experts that have a deep understanding of human attentional processes. The SEEV model relates the probability P_s of attending a specific AOI s to four factors:

$$P_s = \underline{S}aliency - \underline{E}ffort + \underline{E}xpectancy \cdot Task\ \underline{V}alue$$

The first two coefficients, *Saliency* and *Effort* are bottom-up factors that describe the saliency of information displayed by an AOI and the effort it takes to obtain the information, e.g., by moving eyes and head or navigating through a menu. *Expectancy* and *Task Value* are top-down factors. They describe how often new information can be expected from an AOI and how valuable the information is for accomplishing the tasks of the human operator.

SEEV model variants, considering some or all of the four factors, have been used to model and predict attention allocations for a wide variety of tasks in various domains: For instance in aeronautics, to predict optimal scanning paths for landing operations [Wickens et al. (2003)], monitoring while taxing on ground [Wickens et al. (2008)], or the influence of specific cockpit instruments [Goodman et al. (2003)] on monitoring behavior. In the automotive domain the model was applied to evaluate drivers' monitoring behavior while approaching intersections [Bos et al. (2015)] and also to evaluate the influence of secondary tasks [Wortelen et al. (2013), Horrey et al. (2006)]. Recent studies also demonstrate modeling efforts ending with valid predictions for nurses' experience level when assisting in an operation theater in a hospital [Koh et al. (2011)].

Research on model-based attention prediction is being performed since several decades. But to the best of our knowledge, the SEEV model parameter estimation is mostly based on pen-and-paper techniques (like e.g. by sheets and matrixes) summarizing discussions between domain and HF experts. Some use simulation environments e.g. to estimate the bottom up parameters of the SEEV model such as MIDAS [Corker & Smith (1993)] for instance, which is developed by the NASA since 1985 to support 3-D rapid prototyping of human-machine systems, to evaluate procedures, controls and displays before they are actually being built in hardware. It was also integrated into the CASCaS framework, which is used to simulate safety-critical tasks [Wortelen (2014)].

Some tools have been proposed to support cognitive model creation. CogTool [Harris et al. (2010)] for instance supports the generation of ACT-R [Anderson et al. (2004)] models with deterministic sequences of actions. These models are based on GOMS and KLM and are targeted on predicting task performance of Windows-, Icons, Menus, and Pointer (WIMP)-based user interfaces. The Distract-R system by Salvucci [Salvucci (2009)] is also based on ACT-R. It allows to create ACT-R models of in-vehicle, secondary task interactions in a way similar to CogTool. It integrates these models with a detailed driver model to simulate and predict effects of secondary task distraction on driving behavior. COGENT [Cooper & Fox (1998)] is a graphical modeling editor for psychologists that allows "programming" cognitive models at a higher level of abstraction. It is based on box and arrow diagrams that link to a set of standard types of cognitive modules, which implement theoretical constructs from psychological theory. COGENT, CogTool, Distract-R and HEE share the idea of making cognitive modeling easier by allowing programming on a higher level of abstraction. Whereas COGENT focuses on psychologists and extensive training, the HEE, CogTool and (to a lesser extent) Distract-R do not require any specific expertise to generate cognitive models and can therefore be used by non-experts in cognitive modeling as well.

To the best of our knowledge, the application of these models is still performed in a niche, mainly in the safety-critical systems domain. Prior works that we are aware of mainly focus on the validity of attention prediction. We show in the following that the model input validity is the main challenge that prevents broader application. Even though, validity is desirable, model exploration offers an additional major benefit independently of validity. Therefore, this work demonstrates the potential of model exploration for HMI design.

2.1 Model Parameter Input Validity

All SEEV model related studies we referenced above report moderate up to very high correlations ($0.6< R <0.97$) between eye tracking studies and the model predictions based on the SEEV model. But input validity remains a problem: the expert knowledge that is required to determine the actual SEEV model parameters to generate valid predictions is not captured. While the bottom up parameters values can be estimated e.g. based on physiological data about the effort for eye and head movements [Gore et al. (2009)] or by computing saliency maps [Itti & Koch (2001)], the determination of the knowledge-based expectancy and value coefficients often depend on data gained by domain experts for a specific application use case.

Most studies therefore rely on methods and techniques to estimate the value and expectancy coefficients based on expert judgment. The broad majority of the studies above applied the "least integer ordinal value" heuristic, which estimates parameter values by letting experts systematically compare AOIs between conditions. A recent approach applies the analytic hierarchy process technique for quantifying the informational importance [Ha & Seong (2014)].

The results of those methods, the relevant concrete parameter values are stated in most of the studies above and predictions therefore can be reproduced, but only one study we found [Koh et al. (2011)] reported insights about the amount of experts, their background and prior knowledge, and the method applied to agree on the concrete model input parameter values. If the attention model is created for instance by only one HF expert, errors made by this HF expert can have a huge impact on the predictions. If the parameter estimation is a result of a discussion of several experts, quiet voices can be overheard easily. Finally, if instead several experts are individually applying a method, the often observed evaluator effect might become evident. This means that individual rating variability is high [Feuerstack & Wortelen (2016)].

2.2 Model-based Exploration

The focus of research of most of the contributions on attention modeling that we are aware of, is on identifying and assessing the relevant factors and their corresponding interactions to explain an observed human behavior. A benefit that one intuitively might expect from the model-based method is that it is a low cost replacement for early eye-tracking studies. Therefore the way that the predictions of a model-based method are utilized often corresponds with how eye-tracking data is utilized. Both methods can be applied as a tool to measure how a human divides attention between several areas of interest (e.g. between products in a shopping mall). This helps understanding what is relevant for a certain task (e.g. eye tracking studies with users performing tasks on a website). However, we argue that model-based predictions do not fully correspond with eye tracking data, because there are structural and qualitative differences.

There is a trade-off between the validity of the data and the insights it provides. A carefully planned and executed eye-tracking study results in highly valid data, as it directly measures the gaze of the human operator. In contrast the validity of model predictions strongly depends on the level of detail of the model and the process of assessing free parameters. As described in Section 2.1 documenting the process of parameter assessment in a reproducible way is difficult. However, eye tracking studies also have some shortcomings concerning the interpretation of the data:

1. Eye-tracking studies measure focal visual attention. With eye trackers it is very difficult to measure the mental focus of attention or information that is perceived peripherally. This typically requires special study set-ups, like occluding parts of the visual field [Land & Horwood (1995)]. Models on the other hand can capture these aspects.

2. Eye-tracking studies explain where operators look at, but not why. Models on the other hand explain the causal relationship between the influencing factors and the distribution of attention. Models can furthermore distinguish the effect of different influencing factors, like saliency or information value.

We claim that for many applications the model-based approach is better suited for explorative analysis and for developing hypothesis of what drives the allocation of attention in specific scenarios. This helps for example to explore the level of shared understanding within an HMI design team or between the designers of an HMI and the domain experts respectively the HMI users.

The Human Efficiency Evaluator (HEE) is a software tool designed to support HMI developers with a more objective view on human-system interaction in an early design phase. It supports a model-based attention prediction method, and helps to document the assessment of free parameters. In prior experiments we have focused on input parameter validation [Feuerstack & Wortelen (2016)]. Therefore, this contribution is about applying the same tool for a model-based exploration to reveal differences in understandings between a group of collaborating HMI designers or between HMI designers and their targeted audience: the operators. In the following sections we describe the tool and report about the two use cases.

3. HUMAN EFFICIENCY EVALUATOR

The Human Efficiency Evaluator (HEE) has been designed to be applied in an early HMI design phase, in which design work focuses on pen- and paper prototypes or early design sketches. The tool implements a structured process for attention modeling. As initial input the HEE requires a set of images, each depicting a design variant of the HMI embedded into the environment (e.g. a car) and in the same specific situations as the other variants. Additionally, a set of user tasks relevant for monitoring the HMI is pre-set.

In the *first step* of the process all AOIs relevant for the operator tasks are to be identified by the HEE user. An AOI is a location within the HMI or in its surrounding environment, from which information can be perceived by the operator. For this step one has to carefully distinguish between information and the source of the information. Information that is tightly connected to the specific point in time should be abstracted to their source location by answering: “Where do I usually expect these information to appear?”.

Figure 1 shows the first screen mask of the HEE web application, which in this example displays a design sketch of an electronic sea chart with already several AOIs identified by a domain expert (a ship master in this case). The boxes above the HMI sketch list three main monitoring tasks of the ship master. For that he should identify relevant AOIs.

In the *second step* the expectancy coefficient for each AOI is calculated using the least integer ordinal value heuristic [Wickens et al. (2001)]: The user has to sort (“rank”) all previously identified AOIs. Those AOIs are put on top of the list for which the user expects to perceive new information most frequently. This is done by roughly ordering the AOIs using a drag- and-drop technique. The HEE transforms this list into a set of “greater than” relations like shown by Fig. 2. More relations can then be interactively added in a subsequent step, specifically to express differences in the expected frequency of information for the same AOI in the

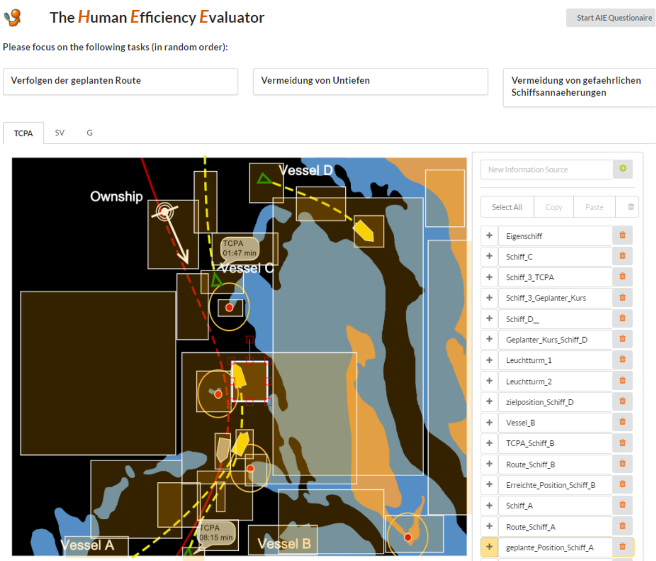

Fig. 1. HEE AOI identification step of the domain expert for the maritime user case: three user tasks are depicted on top.

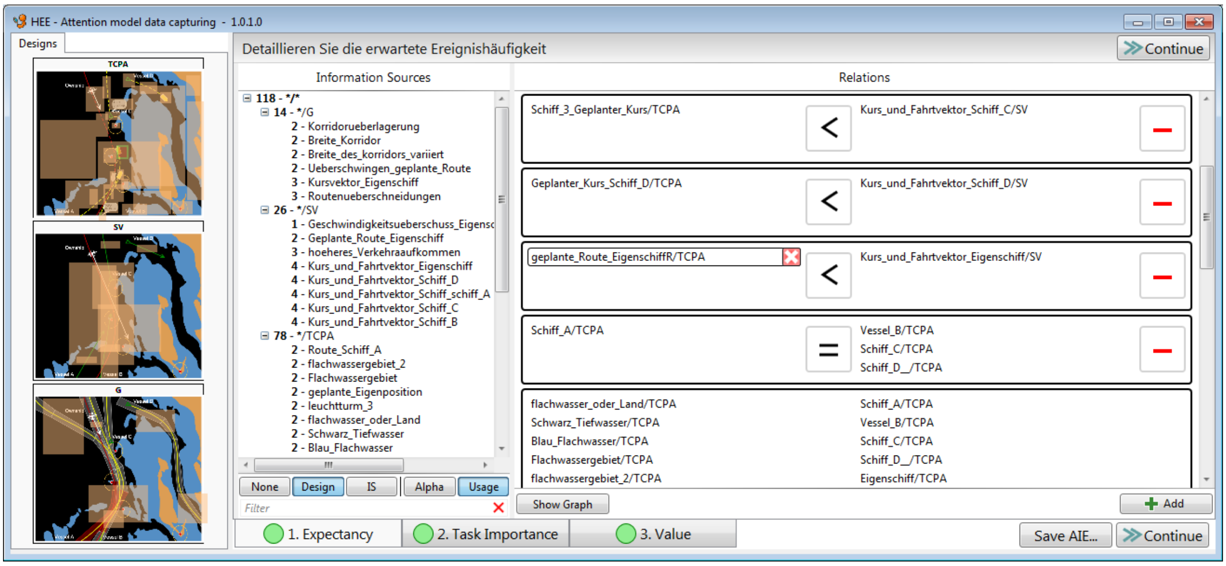

Fig. 2. After roughly sorting the AOIs, greater-than relations are automatically created (to the right). New relations can be created by dragging the relevant AOIs from the list depicted in the center into a new relation at the right.

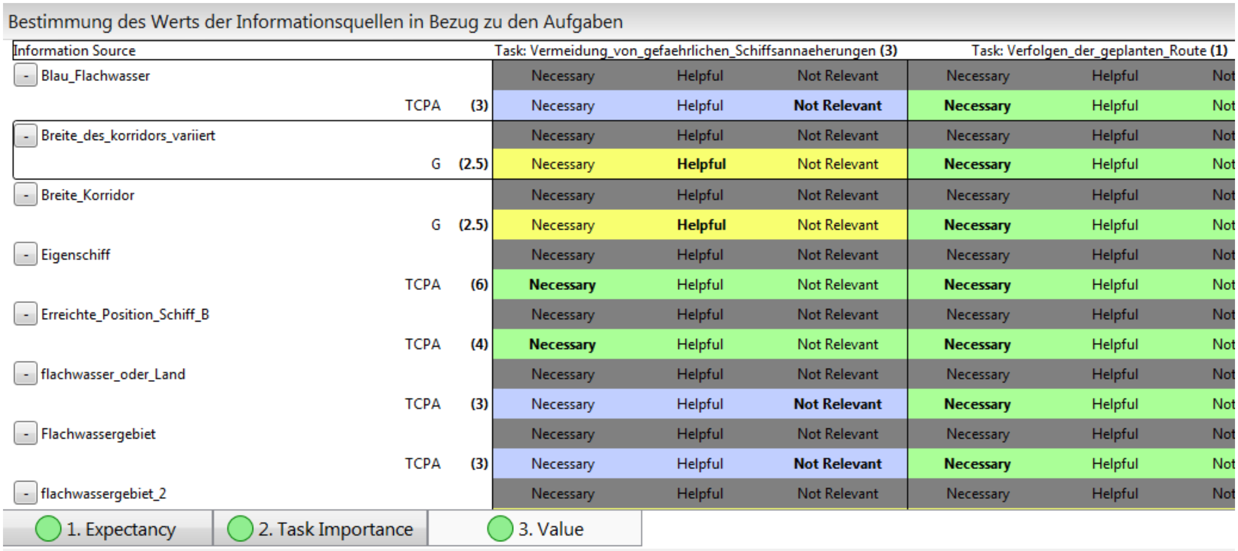

Fig. 4. Excerpt from the relevance matrix. AOIs are listed by rows, tasks are identified by columns. A colour scheme highlights the user's choices between "necessary" (green), "helpful" (yellow) or "not relevant (blue).

different design variants. To support this step, additional relations can be created and AOIs can be placed into the left or right hand side of a new relation using drag-and-drop. The HEE highlights the currently selected AOI graphically on the design and also checks and highlights contradicting relations.

In the *third step*, the operator tasks are ordered by their importance, with the most important tasks on top of the list. Fig. 3 shows the corresponding screen mask of the HEE: tasks listed on the right have already been ranked.

Finally, in the *fourth step*, the relevance of each AOI for each of the operator tasks is determined by completing a relevance matrix, which lists all identified AOIs as rows and the user tasks as columns. The HEE user has to decide if the corresponding AOI is either "required", "helpful" or "not relevant" for the corresponding task. Fig. 4 depicts the screen mask of the HEE for the relevance determination.

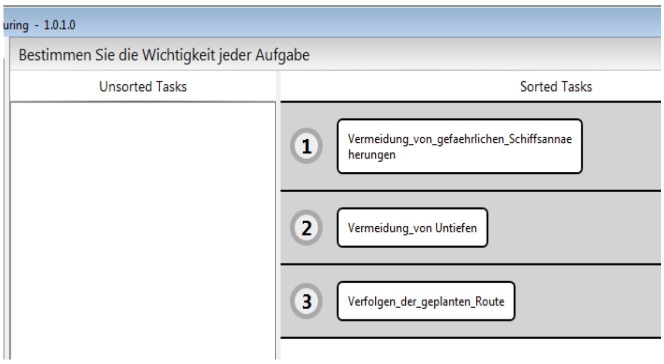

Fig. 3. Each tasks of the unordered task list at the left need to be ordered by dragging it into the list on the right.

Based on this collected data, predictions of visual attention distributions and reaction times to unexpected events can be calculated [Wortelen & Feuerstack (2016)]. An HEE project can be completed independently by several HEE users resulting in several attention models for the same design. Recording multiple models from multiple users and averaging the results improves model validity, because random individual errors are cancelled out. The reproducibility of the modelling process is enhanced by the HEE. The HEE project describes the entire study setup and the process of data acquisition. The recorded data describes the free parameters of the attention model that were identified by the HEE users during this process. Because the study description and the recorded data are stored in a structured data format, the documentation of the modelling process is quite easy.

Besides computing predictions as done by others (c.f. sec. 2), such model-based data gained by several experts can also be explored to discover reasons and hidden effects, which will be discussed in detail in the upcoming sections.

4. MODEL EXPLORATION

There are many methods for eliciting knowledge for designing interactive systems. The user-centered design (UCD) process [ISO (2010)] for instance identifies a set of subsequent phases including specifying context and requirements, prototyping designs, and finally evaluating them with users. Popular UCD methods include participatory design [Schuler & Namioka (1993)] that involves stakeholders to ensure that the design result meet their needs, or interviews, focus groups or questionnaires to identify user-needs and to better understand the requirements.

The better the domain, the user tasks, and the users' knowledge is understood by the designer, the higher is the chance that the HMI design matches users' needs. For most of these methods objectivity is hard to maintain, because they collect subjective data from experts. Some require moderation or other forms of unstructured or semi-structured interviews, which impairs reproducibility. Contrarily, formal questionnaires or pre-structured interviews offer the chance of result-reproducibility but also limit the chance to explore something not known beforehand.

There are other methods available, such as ecological interface design (EID) [Vicente & Rasmussen (1992)] for instance, which does not focus on the user, but on the work-domain from that the constraints of the environment and the objectives of the domain are derived. Approaches concentrating on the work domain have the advantage that they can also discover and consider situations that are unexpected by the users and aim at improving human performance by reducing their workload [Vicente & Rasmussen (1992)]. Recent approaches, like Konect for instance specifically focus on optimizing designs for fast and correct visual perception [Ostendorp et al. (2016)]. Konect embeds heuristics and basic research on pre-attention into a design method to guide design of visual user interfaces that are optimized for fast and correct perception.

Model exploration with the HEE is very rigid in the structure and for the inputs that are collected from the users of the tool. The objective is to collect subjective data from experts in a highly reproducible way.

While most methods require experience and training to be successfully applied, the HEE does not require extensive training. Prior studies have shown that a short 12 minutes video tutorial[1] is sufficient to enable novice users to model their monitoring behavior with the tool [Feuerstack et al. (2016), Feuerstack & Wortelen (2016)].

We experimented in two different settings with model exploration: First, we applied the HEE to compare the view of an interaction designer with the one of a targeted user (a ship master) on three different abstract design sketches of an electronic chart display and information system (ECDIS). Second, we compared the view of six different HF experts on three different automotive HMI versions of a traffic light assistance system.

4.1 Exploring Designer's Perspective vs. User's Perspective

In the maritime domain Electronic Chart Display and Information Systems (ECDIS) are one of the main sources of information that are monitored to support vessel navigation. Often the complete passage is planned and inserted into the vessel navigations system prior to the trip and one of the main tasks of a ship pilot during the trip is to observe the own vessel and to monitor for other vessels and unexpected obstacles. Following a UCD process, an HMI designer had sketched three design alternatives for an improved ECDIS display.

We asked the HMI designer to use the HEE and to model how he assumes that a ship master will monitor the three design variants. Thereafter, we presented and explained the design sketches to a ship master and asked him to model his monitoring behavior for each of the three designs. With this input we simulated the monitoring behavior for both experts and for each design and generated several comparative visualizations that we presented the HMI designer for exploration [Feuerstack & Wortelen (2015)].

Fig. 5 depicts such a set of comparative visualizations for one of the HMI designs. The design alternatives were specifically elaborated to support the ship master in a critical situation. Therefore, we chose a situation in that the own vessel ("ownship" of Fig. 5b+c) was required to pass through a very narrow sea strait with oncoming traffic of two other vessels. The overall design idea of the depicted HMI variant of Fig. 5 (and all subsequent figures of this section) was to highlight predicted vessel routes depicted by dashed yellow lines to improve a ship master's anticipation capabilities of critical situations. The yellow vessel symbols therefor identify future vessel positions. Different to what a non-expert in maritime operations would expect, ECDIS systems color shippable water in black and shallow water conditions (that need to be prevented) in blue. Circles with a central dot represent lighthouses, which identify shallow water in this situation.

[1] http://lnk.multi-access.de/kogsys17, last checked 03/09/17

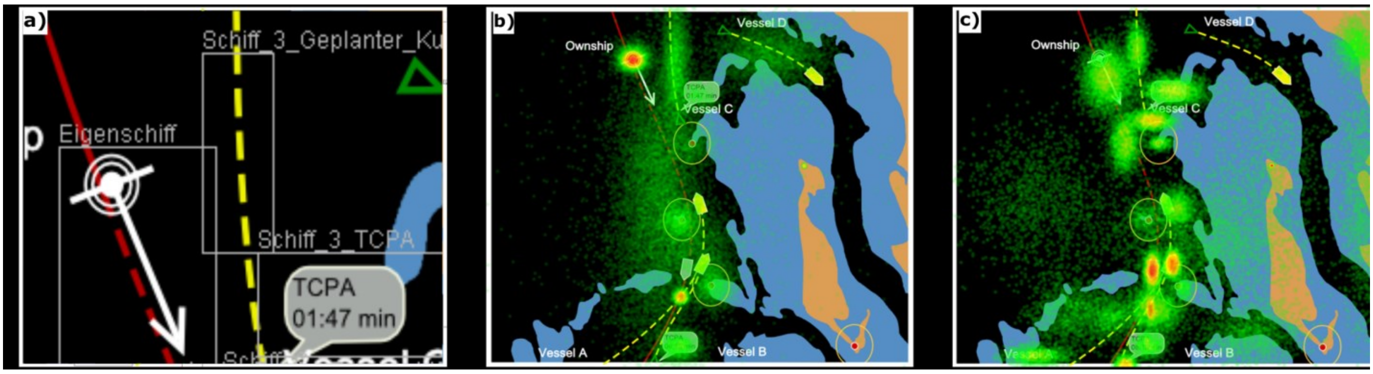

Fig. 5. Heat map visualization. a) The boundaries of the operator's AOIs (thin gray lines) used as reference for the interpretation of the operator's heatmap. Shown is only a small, zoomed-in part of the entire sea chart for better readability. AOI names are in German, because the operator and HMI designer were German; The heat map resulting from the simulation of the cognitive model defined by b) the HMI designer and c) the operator. Simulation was performed by automatically creating a cognitive model in CASCaS using the designer's and operator's input data [*Wortelen et al. (2013)*].

Fig. 5 depicts such a set of comparative visualizations for one of the HMI designs. The design alternatives were specifically elaborated to support the ship master in a critical situation. Therefore, we chose a situation in that the own vessel ("ownship" of Fig. 5b+c) was required to pass through a very narrow sea strait with oncoming traffic of two other vessels. The overall design idea of the depicted HMI variant of Fig. 5 (and all subsequent figures of this section) was to highlight predicted vessel routes depicted by dashed yellow lines to improve a ship master's anticipation capabilities of critical situations. The yellow vessel symbols therefor identify future vessel positions. Different to what a non-expert in maritime operations would expect, ECDIS systems color shippable water in black and shallow water conditions (that need to be prevented) in blue. Circles with a central dot represent lighthouses, which identify shallow water in this situation.

Fig. 5b+c depict the heatmap visualizations, which were generated based on simulating the monitoring behavior of the HMI designer (b) and the ship master (c). To ease comparing both, the HEE also generates images with labeled AOIs (c.f. Fig. 5a). Comparing both heatmaps, it is evident that the designer assumed that the ship master spends much more attention on monitoring the position of his own ship than he actually intents to do. Further on, the ship master identified the direction and not the position of the own ship as a monitoring target and intends to invest much more time on observing the future positions of the vessels and potential overlapping routes than expected by the designer.

It can be further observed already from the heatmap visualizations that the ship master intends to monitor more areas than it was assumed by the designer. By visualizing just the identified AOIs as boxes, like shown by Fig. 6a for the designer and Fig. 6b for the ship master, this observation becomes even more evident: The ship master distributes his visual attention on the entire map, but focusses the attention around the area where the vessels have the closest point of approach. Contrarily, the designer assumed that the focus of the shipmaster is much more centered on the vessels route predictions. The grey level of both visualizations reflects the expectancy level. Those areas with a darker grey level identify areas with a high expectancy of being a source for retrieving often new information. Both expect to perceive frequent new information from the situation around the narrow sea passage, but interestingly the designer seem to overestimate the amount of new information that a ship master expects close to the own vessel.

One can also explore and compare which AOIs are assumed to capture most of the visual attention. Such a visualization is depicted by Fig. 7a for the designer and by Fig. 7b for the ship master. By comparing both, one can observe that the designer seems to overestimate the impact of displaying the planned own ship route and the route predictions of the other vessels. The former seems not relevant for the ship master whereas the latter are essentially relevant for observing a route crossing and the vessel that approaches next.

Finally, differences between individual AOIs can be further explored. Therefore the HEE calculates pairs of the geometrically most similar AOIs between the ones from the designer and the ship master. The geometrical similarity is calculated by the root integrated squared distance, which is sensitive to differences in size and position of two rectangles.

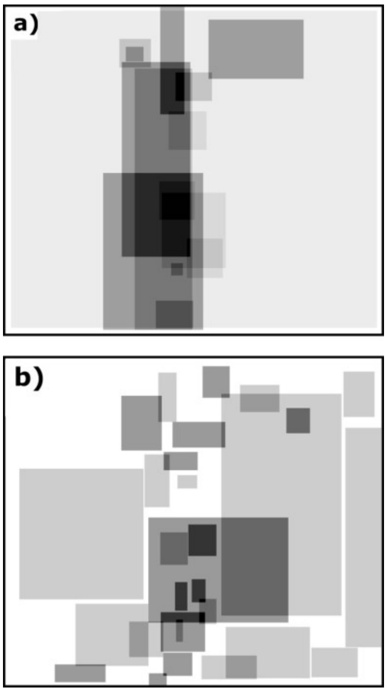

Fig. 6. Colorization of AOIs based on expectancy coefficients defined by (a) the HMI designer and (b) the operator.

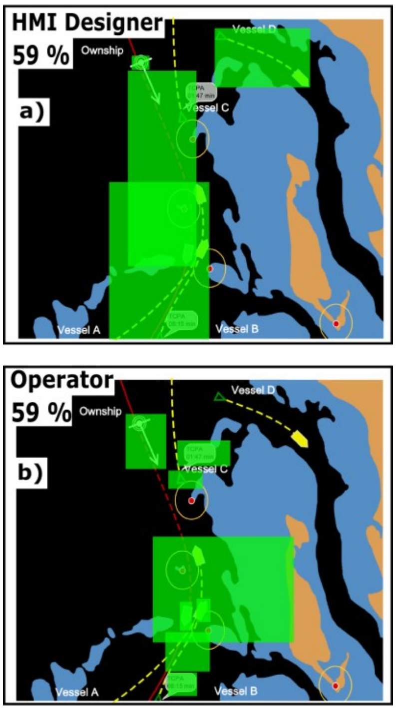

Fig. 7. AOI that attract 59% of the attention as predicted by (a) the designer and (b) the ship master.

Based on this calculation the HEE can display: (1) The most similar AOI pairs with high differences in either expectancy, value or even both, and (2) AOIs defined by the HMI designer, for which only bad matches can be found in the AOIs of the ship master (and vice versa).

Fig. 8 depicts an example for (1) and compares the AOI that both have identified as the "narrow passage"/"high traffic area". The respective expectancy and value ratings from the designer (blue) and the shipmaster (green) are shown below the figure together with the total visual attention that each intends to spend of monitoring the AOI. For this AOI the designer underestimated the amount of new information that this AOI provides for the ship master.

Also bad matches (2) can give interesting insights by supporting the designer to discover unconsidered AOIs, which are in fact relevant for an operator. Fig. 9 shows such an example: For the AOI that was named "possibility of crossing traffic" by the shipmaster the best match that was found is pretty bad. It is a light house and has nothing to do with the AOI defined by the ship master. The existence of crossing traffic was not considered by the designer as its identification requires experience in analyzing the topology of the visualized map (the ship master inferred a close but non-displayed port in this case).

4.2 Exploring a Design Group's Understanding of an HMI

Instead of confronting a designer with the perspective of a later user another use case of the HEE is to explore the potential different understandings within a team of designers working on HMI design proposals. In discussions within the team quite voices can be overheard easily and in bigger teams participants might also talking past each other without even noticing.

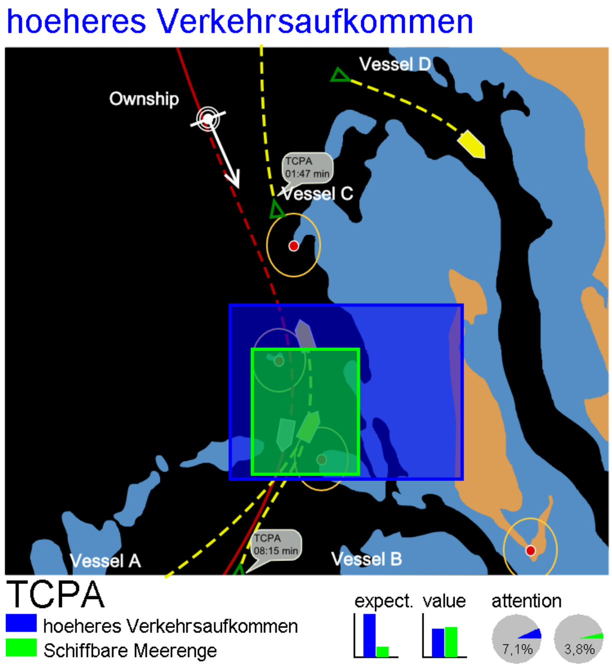

Fig. 8. Direct comparison of two AOIs with high geometric similarity: from designer (green) and ship master (blue).

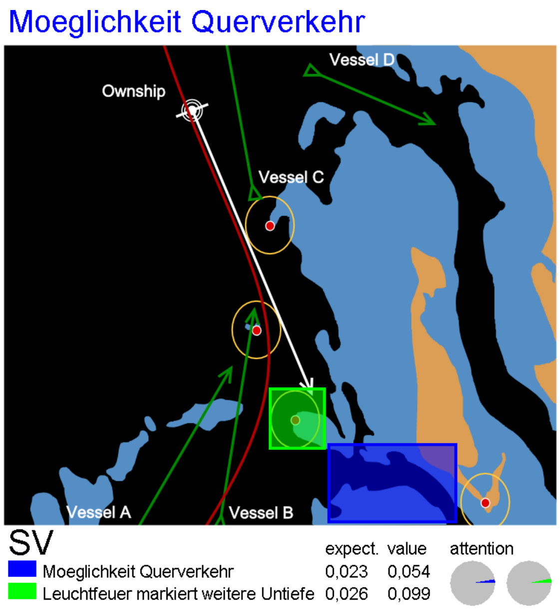

Fig. 9. Direct comparison of two AOIs with very low geometric similarity: from designer (green) and ship master (blue).

The HEE can be applied to discover different perspectives between the designers enabling them to align each other and to discover potential inconsistent views. In the following we present a use case in that an automotive HMI was being evaluated by a total of six Human Factors and Design Experts.

Fig. 10 depicts three different visual HMI variants of an Urban Automatic Cruise Control (ACC) system, which was designed to support drivers to better understand their vehicles' automatic speed adjustments [Kettwich et al. (2016)]. Equipped with an Urban ACC and vehicle-to-infrastructure communication the vehicle is able to adapt its maneuvers with regard to the traffic light signal status and its

Fig. 10 Display structures of three HMI variants; HMI1: Pictogram of a traffic light with radio waves; HMI2: Current traffic light signal status with countdown; HMI3: Traffic light signal status when arriving and passing the corresponding intersection

future phase change to increase traffic safety and traffic flow efficiency. There are situations in that the vehicle reacts to a forecast, which might be in discrepancy with the current environmental situation (e.g. traffic light "green", vehicle is decelerating) and the three visual HMI variants were aimed at improving a drivers understanding of the Urban ACC behavior.

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.

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 Fig. 11). Four tasks that represent typical urban traffic driver tasks have been predefined to be considered for modeling the visual attention with the tool:

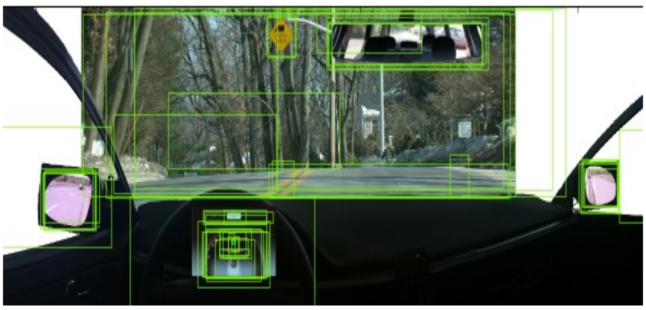

Fig.11 Sign that announces upcoming traffic light. Green boxes show areas of interest identified by HF experts.

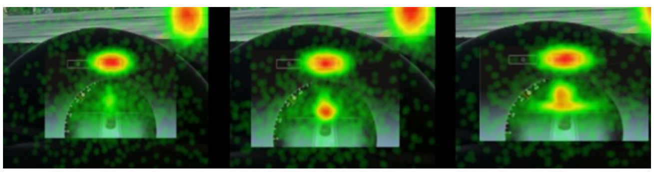

Fig.12 Visual attention prediction for the HM1-3 variants of one HF expert.

"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 functionality

The six HF experts applied the HEE process independently from each other on their own computers and submitted the data thereafter to us for analysis. Fig. 11 depicts in green boxes the AOIs identified by all subjects for the HMI3 variant. One can see that they identified mostly the same areas. Heatmaps with visual attention predictions can then be calculated with the HEE. Fig. 12 for instance visualizes shifts of visual attention between the HMI variants based on data gained by one HF expert. If there is a shift in attention dependent on the HMI variant, it is of interest to analyze, whether it is predicted by all experts and where the attention is drawn from.

To explore the differences between the HF experts by their value and expectancy ratings, we manually classified all AOIs with three raters into 15 classes (with a high inter-rater reliability: Fleiss' $\kappa = 0.88$). In general, all ratings for the tasks importance and expectancy rankings and also for the relevance matrix from all HF experts were highly concordant (Kendall's coefficient of concordance $Wt > 0.82$; $p < 0.01$) for those AOIs that all HF experts have identified: The forward view, the Urban ACC HMI, and the speedometer.

The left graph of Fig. 13 shows the mean percentage dwell time (PDT) for the ACC HMI and the forward view. We found a significant difference of predicted attention distribution between the HMI variants for the Urban ACC display ($F(2,10) = 8.041$, $p = 0.00828$). Subsequent t-tests indicated, that the mean PDT for HMI 1 ($M=0.069$, $\sigma=0.067$) was significantly different to HMI 3 ($M=0.106$, $\sigma=0.072$, $p=0.002$) and also to HMI 2 ($M=0.113$, $\sigma=0.063$), 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 ($F(2,6)=7.054$, $p=0.027$).

The same kind of PDT analysis can also be done with data from eye-tracking studies. But at this point the opportunities of the model-based approach emerge. The model-based approach of the HEE allows inspecting the simulation models to learn about causes for the observed shifts in attention between the HMI variants and to identify whether all experts assume the same causes for shifts in attention.

By looking at the expectancy, relevance and task importance coefficients one can explore causes for this effect. For this

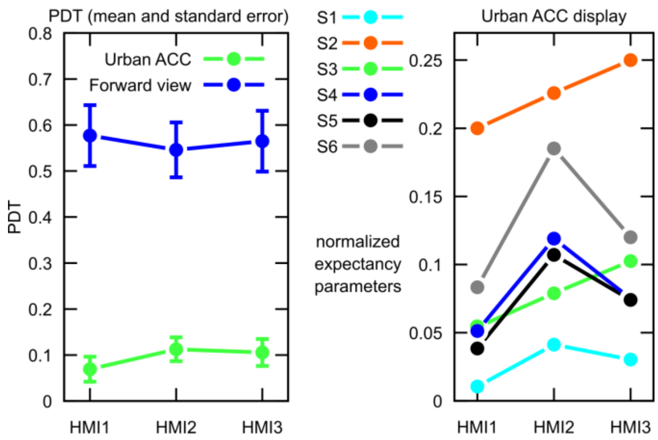

Fig. 13. Left: percentage dwell time. Right: normalized expectancy parameters of ACC HMI for all HF Experts

case, actually neither the task importance parameters, nor the relevance matrix, but the difference in the amount of expected information from the Urban ACC display explained the effect. The standard deviation of expectancy parameters for the Urban ACC display (HMI 1=0.067, HMI 2=0.068, HMI 3=0.076) was by far higher than for any other AOI. The right graph of Fig. 13 depicts the normalized expectancy parameters for the Urban ACC HMI of all six HF experts. It can be observed that (1) they all expect a different amount of new information from each HMI variant and (2) also their opinions differ: All expect the least information from HMI 1, but experts S2 and S3 expect the most information from HMI 3, while all other experts expect it from HMI 2. Searching for such patterns can be extremely helpful information for an HMI designer to check whether the own expectation about the HMI design matches with the expectations of the others. Such inconsistencies can be a starting point for discussion in the design team.

Similar patterns can be found in the relevance definition (even though experts rated highly concordant as we already mentioned earlier): As depicted in Fig. 14 the HF experts disagreed about the relevance of the HMI variants for the "Control Speed" task, which is in fact an interesting aspect to initiate a discussion between the experts.

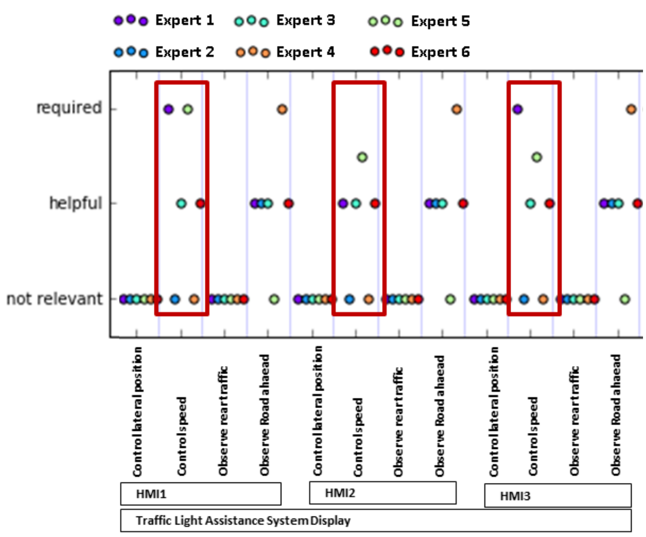

Fig.14 Relevance ratings of the HF experts for each HMI variant for the ControlSpeed driver task.

Finally, for this use case the experts generally agreed about that the "Control Lateral Position" and "Observe Road Ahead" tasks are more important than the "Control Speed" and "Observe Rear Traffic" tasks.

5. CONCLUSIONS

The Human Efficiency Evaluator can be used in a very efficient manner to generate monitoring behaviour predictions and can be applied complementary to eye tracking studies. Whereas the latter provides an objective measurement but requires a working prototype, our approach benefits from the ability of humans to reflect about their mental model of a situation or design based on something as simple as a sketch or a photo [Feuerstack & Wortelen (2015)].

Whereas eye tracking studies typically require a physical or simulated functional HMI prototype and usually depend on serial subject processing to collect monitoring behaviour data, the HEE can be used remotely, in parallel sessions, without any specific hardware, with only minor training, and can also be applied to analyse early designs for that no functional prototype is available. The tool aggregates the collected data, automatically generates a cognitive model that simulates monitoring [Feuerstack & Wortelen (2016)] of the HMI designs, and results in visualizations that summarize the simulation data for further analysis. Eye-tracking data per se gives no explanation of the behaviour recorded, whereas a model based simulation can be inspected e.g. for hidden effects and reasons for the simulated behaviour.

We demonstrated by two exemplary use cases how these visualizations can support HMI designers and HF experts in analysing their HMI designs and exploring the targeted users monitoring behaviour.

In a maritime use case we generated comparative visualizations of two monitoring behaviour predictions. One prediction was generated by the HMI designer who designed the HMI and the second prediction was generated by a ship master – the targeted user. It turned out that the comparative visualizations helped the HMI designer to view the design with the eyes of the ship master. The designer afterwards better understood what the ship master focuses on the most and why.

In a second use case we evaluated the impact of three design variants of an automotive HMI assistance system [Feuerstack et al. (2016)] for the monitoring behaviour with six HF experts. This study revealed that between experts there are differences in how they expect the car drivers to use and monitor the designs.

The HEE fosters a very fast and easy to use data capturing and monitoring behaviour modelling even for non-experts in cognitive modelling. In the studies we performed, none of the subjects required more than a total of 90 minutes to understand the tool usage (by watching a 15 minutes video tutorial), and to end up with a monitoring behaviour prediction.

Optimizing HMIs for efficient monitoring reduces reaction time on unexpected events and improves situation awareness of the user. Both are very relevant especially for safety-critical application. Comparative visualizations of monitoring behaviour can already discover misunderstandings and misleading assumptions between the HMI designers in the HMI development team or between HMI designer and the end user in an early design phase.

ACKNOWLEDGMENT

The authors acknowledge the financial support by the European Commission (H2020-MG-2014-2015) in the interest of the project AutoMate (http://www.automate-project.eu/) – Grant Agreement 690705 and the funding initiative Niedersächsisches Vorab of the Volkswagen Foundation and the Ministry of Science and Culture of Lower Saxony as a part of the Interdisciplinary Research Centre on Critical Systems Engineering for Socio-Technical Systems.

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C. & Qin, Y. (2004), 'An integrated theory of the mind.', *Psychological Review* **111**(4), 1036–1060.

Bos, A. J., Ruscio, D., Cassavaugh, N. D., Lach, J., Gunaratne, P. & Backs, R. W. (2015), Comparison of novice and experienced drivers using the SEEV model to predict attention allocation at intersections during simulated driving, *in* 'Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design'.

Cooper, R. & Fox, J. (1998), 'Cogent: A visual design environment for cognitive modeling', *Behavior Research Methods, Instruments, & Computers* **30**(4), 553–564.

Corker, K. M. & Smith, B. R. (1993), An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation automation, *in* 'Proceedings of the AIAA Computing in Aerospace 9 Conference', pp. 1079–1088.

Feuerstack, S., Lüdtke, A. & Osterloh, J.-P. (2015), A tool for easing the cognitive analysis of design prototypes of aircraft cockpit instruments, *in* 'Proceedings 33rd annual conference of the European Association of Cognitive Ergonomics', ACM, pp. 22:1–22:8.

Feuerstack, S. & Wortelen, B. (2015), Revealing differences in designers' and users' perspectives: A tool-supported process for visual attention prediction for designing HMIs for maritime monitoring tasks, *in* 'Proceedings of INTERACT 2015', LNCS 9299, Springer, pp. 105–122.

Feuerstack, S. & Wortelen, B. (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.

Feuerstack, S., Wortelen, B., Kettwich, C. & Schieben, A. (2016), Theater-system technique and model-based attention prediction for the early automotive hmi design evaluation, *in* P. Green, S. Boll, G. Burnett, J. Gabbard & S. Osswald, eds, 'Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications', ACM, New York, USA, pp. 19–22.

Goodman, A., L.Hooey, B., Foyle, D. C. & Wilson, J. R. (2003), Characterizing visual performance during approach and landing with and without a synthetic vision display: A part task study., *in* D. C. Foyle, A. Goodman & B. L. Hooey, eds, 'Proceedings of the 2003 Conference on Human Performance Modeling of Approach and Landing with Augmented Displays', number NASA/CP-2003-212267, Moffett Field, CA: NASA, pp. 71–89.

Gore, B. F., Hooey, B. L., Wickens, C. D. & Scott-Nash, S. (2009), A computational implementation of a human attention guiding mechanism in midas v5, *in* V. G. Duffy, ed., 'Digital Human Modeling', Vol. 5620/2009 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 237–246.

Ha, J. S. & Seong, P. H. (2014), 'Experimental investigation between attentional-resource effectiveness and perception and diagnosis in nuclear power plants', *Nuclear Engineering and Design* **278**, 758–772.

Harris, B. N., John, B. E. & Brezin, J. (2010), Human performance modeling for all: Importing ui prototypes into cogtool, *in* 'CHI '10 Extended Abstracts on Human Factors in Computing Systems', CHI EA '10, ACM, New York, NY, USA, pp. 3481–3486.

Horrey, W. J., Wickens, C. D. & Consalus, K. P. (2006), 'Modeling drivers' visual attention allocation while interacting with in-vehicle technologies.', *Journal of Experimental Psychology: Applied* **12**(2), 67–78.

ISO (2010), 'ISO 9241-210:2010 - Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems', International Organization for Standardization, Geneva, Switzerland.

Itti, L. & Koch, C. (2001), Computational modelling of visual attention, *Nature Reviews Neuroscience* **2**(3), 194–203.

Kettwich, C., Haus, R., Temme, G. & Schieben, A. (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)'.

Koh, R. Y. I., Park, T., Wickens, C. D., Ong, L. T. & Chia, S. N. (2011), 'Differences in attentional strategies by novice and experienced operating theatre scrub nurses', *Journal of Experimental Psychology: Applied* **17**(3), 233–246.

Land, M. F. & Horwood, J. (1995), 'Which parts of the road guide steering?', *Nature* **377**, 339–340.

McCarley, J. S., Wickens, C. D., Goh, J. & Horrey, W. J. (2002), A computational model of attention/situation awareness, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', Vol. 46, SAGE Publications, pp. 1669–1673.

Ostendorp, M.-C., Feuerstack, S., Friedrichs, T. & Lüdtke, A. (2016), Engineering automotive HMIs that are optimized for correct and fast perception, *in* 'ACM SIGCHI Symposium on Engineering Interactive Computing Systems 2016', Brussels, Belgium.

Salvucci, D. D. (2009), 'Rapid prototyping and evaluation of in-vehicle interfaces', *ACM Transactions on Computer-Human Interaction* **16**(2), 9:1–9:33.

Schuler, D. & Namioka, A., eds (1993), *Participatory Design: Principles and Practices*, L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Vicente, K. & Rasmussen, J. (1992), 'Ecological interface design: theoretical foundations', *Systems, Man and Cybernetics, IEEE Transactions on* **22**(4), 589–606.

Wickens, C. D., Helleberg, J., Goh, J., Xu, X. & Horrey, W.J. (2001), Pilot task management: Testing an attentional expected value model of visual scanning, Technical report, NASA Ames Research Center Moffett Field, CA.

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M. & Zheng, S. (2008), *Attention-Situation Awareness (A-SA) Model of Pilot Error*, CRC Press/Taylor & Francis Group, chapter 9, pp. 213–239.

Wickens, C., McCarley, J. & Thomas, L. (2003), Attention-situation awareness (A-SA) model, *in* 'NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays', p. 189.

Wortelen, B. (2014), Das Adaptive-Information-Expectancy-Modell zur Aufmerksamkeits-simulation eines kognitiven Fahrermodells, PhD thesis, Universität Oldenburg - Department für Informatik. http://oops.uni-oldenburg.de/1970/1/ worada14.pdf

Wortelen, B., Baumann, M. & Lüdtke, A. (2013), 'Dynamic simulation and prediction of drivers' attention distribution', *Transportation research part F: traffic psychology and behaviour* **21**, 278–294.

Wortelen, B. & Feuerstack, S. (2016), Tool-supported comparative visualizations to reveal the difference between 'what has been designed' and 'how it is perceived' for monitoring interface design, *in* '2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)', San Diego, USA, pp. 199–204.