Revealing Differences in Designers' and Users' Perspectives:

A Tool-supported Process for Visual Attention Prediction for Designing HMIs for Maritime Monitoring Tasks

Sebastian Feuerstack, and Bertram Wortelen

OFFIS – Institute for Information Technology
Escherweg 2, 26121 Oldenburg, Germany
{Feuerstack, Wortelen}@offis.de

Abstract. Monitoring complex systems includes scanning, aggregating and processing data from various sources. The design of graphical interfaces for monitoring tasks involves a fine-grained exploration of the importance and expected frequency of events that an operator needs to be informed about.
The Human Efficiency Evaluator is a tool for the prediction of human behavior. We extended it to predict the distribution of operator's attention while monitoring interfaces. The prediction is based on the SEEV model. We show that our tool can be used by experts with different backgrounds to generate predictions following a structured, semi-automated process.
In a qualitative study with subject matter experts, we analyzed different HMI designs for a navigation task in the maritime domain. We evaluated their modeling time, tested different prediction result visualizations, and investigated in the model differences between the subjects. Different to what we originally expected, the study revealed that the models created by the subjects substantially differ depending on their perspectives. Heat maps visualizing the predicted attention allocation were appreciated by the subjects and enabled them to argue about their perspective.

Keywords. Visual attention, HMI analysis, Monitoring task

1 Introduction

Controlling safety-critical systems often includes several monitoring tasks to observe the current system state and to predict future system states that might affect the controlling activities. A ship bridge is an example for such a safety-critical system. Modern bridge systems offer a broad range of automation of routine tasks and most of the navigation decisions (passage planning for instance) can be performed prior to the voyage. Therefore most of the time on a ship bridge is spent on the navigation monitoring task. This includes observing the ships status, its navigation path and watching out for future events that require adjusting the planned route of the vessel's by changing its speed and heading to prevent dangerous situations.

Studies show that in between 75%-80% of accidents in ship navigation happen because the human operator had not access to information that could have prevented the accident [1]. A study that investigated the lack of situation awareness of mariners revealed the importance of situation awareness for the decision making process in the maritime domain. From the 177 maritime accident reports analyzed, 71% percent of the human errors were situation awareness related. 58.5% of those could be classified as caused by failures in correctly perceiving information [2]. These types of failures are caused by data not being available, data not being easy to discriminate, misperceptions or failures in monitoring or observing data [2].

Electronic integrated bridge concepts are driving future navigation system planning [3]. Such systems aggregate data from various ship sensors and support the vessel navigation task with a map-centric view augmented with current and future navigation paths of the own vessel and real-time information of other vessels in the current area of the ship [3]. Monitoring such complex systems demands intermittently for the human operator's attention to observe and interpret several information sources [4].

This contribution presents a tool, the Human Efficiency Evaluator (HEE), which supports subject matter experts (SMEs), such as HMI designers and domain experts, with no background in cognitive modeling to generate and benefit from attention predictions. With the HEE predictions can be performed by a cognitive human operator simulation in a very early design phase where only HMI design images or sketches of future interfaces are available. The HEE uses the Adaptive Information Expectancy model (AIE model) [5], which is based on the SEEV model [6] and is a dynamic simulation model of attention distribution.

The following section discusses related work. Thereafter we detail the theoretical background of our approach, the process and the underlying models in section 3. Section 4 describes how the process is implemented in the HEE. Section 5 illustrates a use case in the maritime domain that we used to evaluate our approach. This qualitative study with SMEs in section 6 reports about the evaluation of the HEE, which was focused on analyzing the modeling differences between different potential users of the HEE: domain experts, HMI designers, cognitive modelers, and situation awareness experts. Different to what we originally expected, we found much more differences in modeling than similarities between the different SMEs. We close by discussing and summarizing the study results and elaborate a list of hypothesizes derived from observations made during the study. Finally, section 8 concludes and states future work.

2 Related Work

The Adaptive Information Expectancy (AIE) model is a predictive simulation model of attention distribution [7,8]. It is an integral part of the CASCaS (Cognitive Architecture for Safety Critical Task Simulation) architecture [9]. It simulates the attention distribution of a human operator based on two sets of input parameters: *Expectancy* and *Task Value*. They describe how often new information can be expected from an information source (IS) and how valuable the information is for accomplishing the tasks of the human operator. These factors are relevant for showing optimal monitoring behavior and have been shown to be the main drivers for skilled operators like pilots and drivers [10]. An expectancy coefficient u_g and a value coefficient v_g is assigned to each goal g of a cognitive model executed in CASCaS.

A cognitive architecture like CASCaS can be understood as a generic interpreter that executes formalized procedures of a human operator in a psychological plausible way. An overview of cognitive computational models like ACT-R, MIDAS and others is provided in [11, 12]. Computational cognitive models have got the potential to automate parts of human factor analyses during system design. In order to leverage this potential the models have to be embedded in a design tool which can be readily applied by design experts.

Monitoring involves detecting and reacting to events and is composed of a set of monitoring goals. To execute such a goal the human operator looks to the IS that can signal the event. Upon event detection the operator utilizes the perceived information to react to this event. If no event is detected, the operator's attention shifts to another monitoring goal probabilistically based on the expectancy and value coefficients. The probability of switching to goal g among a set G of monitoring goals is defined as (cf. [7]):

$$P(g) = \frac{u_g}{\sum_{g_i \in G} u_{g_i}} \cdot \frac{v_g}{\sum_{g_i \in G} v_{g_i}}$$

Several tools already support cognitive model creation. CogTool [13] supports the generation of ACT-R [14] models implementing deterministic sequences of actions. These models are based on GOMS and KLM and therefore CogTool targets on evaluating Windows-, Icons, Menus, and Pointer (WIMP) user interfaces but also considers speech-commands and basic gestures. For monitoring simulations the HEE generates a probabilistic sequence of actions. Based on CASCaS and the AIE model prediction of average values (percentage dwell times, gaze frequencies) and prediction of distributions (reaction times, duration of diversion) are possible.

MIDAS is a system developed by NASA [15], which uses cognitive models to analyze tasks and interfaces in the aerospace domain. It simulates visual attention distribution similar to CASCaS, but uses the SEEV model of Wickens et al.[6] instead of the AIE model. Both models are strongly related, but differ in some aspects [7].

3 Approach

Cognitive models require input from an expert to predict the attention. In our case these inputs are the identification and location of ISs on a graphical display. Because we use the AIE model (see Section 2) for prediction of attention distribution, for each Information Source (IS) an expectancy coefficient and a value coefficient need to be defined by an expert. Fig. 1 illustrates this. We explain the theory of how to derive the expectancy and value coefficients from the experts' inputs and formalize it for both in the two upcoming subsections. Based on these information and the identification of one IS that most probably make the expert aware of a specific unexpected event (subsection 3.4) the cognitive operator model (subsection 3.5) and the HMI model are generated. After the theoretical foundation, section 4 explains how forms and visuali-

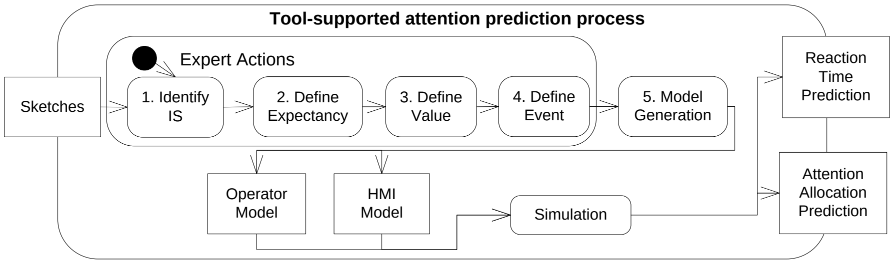


Fig. 1. Main activities of the process to generate the models that are then feed into a simulation to predict reaction times and the attention allocation of the operator.

zations support the coefficient derivation, which is then feed into the Human Efficiency Evaluator for running the simulation and generating the predictions.

3.1 Identification of Information Sources

Information sources (IS) are regions of a display or on a design sketch that communicate a single piece of information to the operator. The process requires that an expert collects all information that can be extracted from a graphical display. Each information is marked as an IS by specifying as exactly as possible its graphical position, its size and by unambiguously naming the IS so that it can be referred to by its name.

3.2 Definition of Expectancy Coefficients

Earlier AIE model applications extracted the expectancy coefficients during the simulation of the cognitive model in interaction with realistic environment simulations [5], [8]. Since this approach is focused on evaluating design sketches, we derive them manually using the lowest ordinal heuristic for this purpose as proposed by Wickens et al. [16]. For this heuristic the expert needs to order all IS based on the amount of new information expected. The rank of an IS in this order determines its expectancy coefficient. The heuristic does not require a total order on the IS. It is sufficient to specify a partial order between the IS across all designs. Partial orders are typically visualized by Hasse diagrams. Following, we formally describe the notation that the expert can use to define the partial order. Let D be the set of all considered designs, S be the set of all defined IS, and $S_D: D \rightarrow \mathcal{P}(S)$ the function that gives for each design the IS that are defined on it. The same IS can be defined on several designs. The tuple $(s|d)$ specifies the IS s defined on design d. By defining relations between different IS such a partial order is established. Relations are formally defined by statements, such as $(s_i|d_j) > (s_k|d_l)$. Which states that IS s_i of design d_j provides relevant events with a higher frequency than IS s_k of design d_l. Such statements can relate IS by the operators “>”, ”<”, or “=”. Further on, several statements can be condensed into a single one to specify relations time efficiently by comma separating IS and designs and also by using quantors. To support an automatic transformation of the relations with quantors into a partial order (visualized as a Hasse diagram) we specified them formally.

Comma separated IS.
Several comma-separated IS and/or designs can be listed on the left hand side and/or right hand side of the relation. Such statements create a set of relationships by using the Cartesian product of the entries on the left hand side and on the right hand side of the statement, e.g.:

$$(s_1, s_2|d) > (s_3, s_4|d) \equiv \left\{(s_i|d) > (s_k|d) \;\middle|\; s_i \in \{s_1, s_2\}, s_k \in \{s_3, s_4\}\right\}$$
$$\equiv \{(s_1|d) > (s_3|d), (s_1|d) > (s_4|d), (s_2|d) > (s_3|d), (s_2|d) > (s_4|d)\}$$

Quantors .
An Asterix is used as quantor to describe that a relation holds for all IS of a specific design $(*\,|d)$ or for a specific IS in all designs $(s|\,*)$. The exact semantic depends on whether a quantor is used on both sides of the relation. Quantor on both sides means Cartesian product of all IS respectively designs. The following specifies that s_1 provides information more frequently than s_2 regardless of a specific design.

$$(s_1|\,*) > (s_2|\,*) \equiv \left\{(s_1|d_i) > (s_2|d_k) \;\middle|\; d_i, d_k \in D\right\}$$

If only one side contains a quantor, it stands for all information sources or designs, except the ones listed on the respective other side. The following specifies that IS s_1 on design d provides information more frequently than all other IS on that design.

$$(s_1|d) > (*\,|d) \equiv \left\{(s_1|d) > (s|d) \;\middle|\; s \in S_D(d) \backslash s_1\right\}$$

A variable following a quantor can be used to bind two quantors on both sides to the same variable. It prevents that the full Cartesian product is created. The following specifies that for each design s_1 provides information more frequently than s_2, but this relation cannot be drawn across different designs.

$$(s_1|\,*Y) > (s_2|\,*Y) \equiv \left\{(s_1|d) > (s_2|d) \;\middle|\; d \in D\right\}$$

The expectancy coefficient for $(s|d)$ is determined by the order of $(s|d)$ within the partial order. In the Hasse diagram of the relation this is the greatest length of a path from any minimal element to $(s|d)$ plus 1. More formally: Let $\leq$ be the transitive reduction of the defined partial order and $\Delta_{\leq}$ be the distance between two elements:

$$\text{expectancy}(s, d) = \max_{d_k \in D, s_i \in S_D(d_k)} \left(\Delta_{\leq}\big((s_i|d_k), (s|d)\big)\right) + 1$$

3.3 Definition of Value Coefficients

Monitoring is typically performed to collect information relevant for tasks a human operator has to perform. The value of an IS depends on the values of the tasks for which it provides information. To obtain the value coefficients for the IS, in a first step the importance of all human operator tasks is rated by the lowest ordinal algorithms as shown above for the expectancy coefficient. Second, the user fills out a relevance matrix R specifying the relevancy R for each IS and each task (cf. [6]). $R(s,t) = 0$ states, that IS s is not relevant for task t, $R(s,t) = 0.5$ defines that s supports t, but is not mandatory for t, and $R(s,t) = 1$ means, that s is mandatory for t. The value of an IS s is determined by (cf. [6]):

$$\mathrm{value}(s) = \sum_{t \in T} R(s,t) \cdot \mathrm{value}(t)$$

3.4 Event Definition

Monitoring is performed in order to maintain suitable situation awareness and to detect events in a timely manner and to react to events if necessary. The reaction time depends on the amount of attention directed to the IS that displays the event. When simulating the allocation of attention, the reaction time to events can also be predicted. An event $e = (s,t)$ is defined by the user by specifying the time t of occurrence and the IS s, that most likely will make the operator aware of the occurrence of this specific event. This might vary between designs. Therefor the IS has to be specified for each design.

3.5 Generation of Monitoring Operator Model

The operator model specifying the monitoring is automatically created based on the IS markups and the expectancy and the value coefficients. For the overall monitoring, a top level goal g_t is created, which has only one associated procedural rule. This rule activates a set of sub-goals – one for monitoring each IS. This structure is reflected by rule R1 in Fig. 2. The sub-goals are annotated with the expectancy and value coefficients as defined above.

```
R1
rule(goal=g_m) {
⇒
  Goal(name=monitor_s_1, value=value(s_1), expectancy=expectancy(s_1, d))
  Goal(name=monitor_s_2, value=value(s_2), expectancy= expectancy(s_2, d))
  …
}
```

```
R2
rule(goal=monitor_s_i) {
  Condition(s_i.event == 'Event')
⇒
  <reaction>
}
```

```
R3
rule(goal=monitor_s_i) {
  Condition(s_i.event != 'Event')
⇒
  --
}
```

```
R4
rule(goal=monitor_s_i) {
  Condition(s_i.event  unknown)
⇒
  LookAt(s_i)
}
```

Fig. 2. Generic structure of rules for simulation of the operator's monitoring behavior in CASCaS-like pseudo code.

Rule R2 is executed, if an event is shown by IS s_i. If no event is shown, R3 is executed. However, as the operator usually does not know the current state of information on s_i the conditions of R2 and R3 cannot be evaluated, because this information is unknown. This state triggers rule R4, which commands the perception and motor component of the cognitive architecture to move the gaze to s_i. After the information about the event is perceived, the model either triggers R2, in case there is an event, or triggers R3, if there is none.

The resulting model, will constantly switch between the monitoring sub-goals to move visual attention between the IS based on the probability calculated with the expectancy and value coefficients.

4 Tool-supported Attention Prediction with the HEE

The Human Efficiency Evaluator (HEE) is based on CogTool because of two reasons. First, we share the idea of supporting an evaluation of interfaces based by annotating design sketches at a very early stage. Second, the results are also predictions based on a simulation.

But there are several fundamental differences between both. CogTool focuses on performance prediction of WIMP (Windows, Icons, Menus, Pointer) interfaces. Therefore, the annotation of design sketches is based on a fixed palette of the most common WIMP widgets like buttons, menu bars, and radio buttons. The annotation process to construct a user interface model is therefore straight-forward by identifying and marking widgets exactly as they are depicted in the design sketch. However, nautical maps have no fixed widgets (even though there are standards, e.g. color) and new design proposals often intentionally break with some of already existing concepts. As we show later, the identification and markup of IS substantially depends on the background and experience of the tool user and also often on interpretations. With CogTool one annotates what is depicted, with HEE one also marks what one knows.

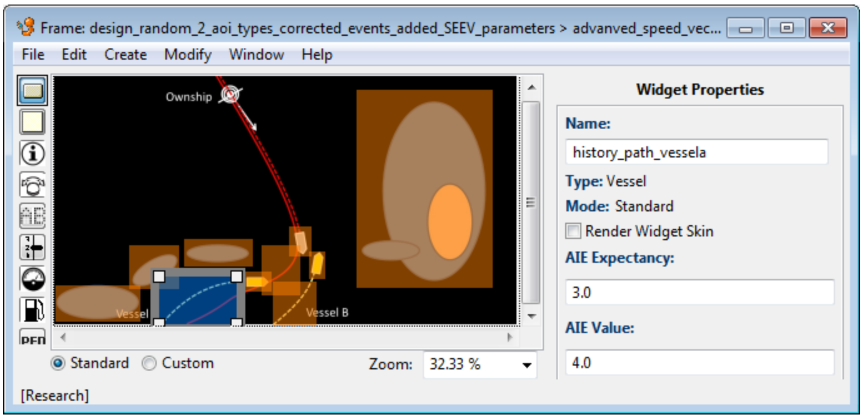
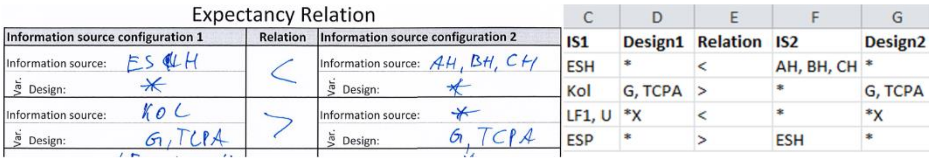


Fig. 3. Design Modeling Perspective of the HEE tool.


Expectancy Relation

Information source configuration 1		Relation	Information source configuration 2	
Information source:	ESH	<	Information source:	AH, BH, CH
Var. Design:	*		Var. Design:	*
Information source:	KoL	>	Information source:	*
Var. Design:	G, TCPA		Var. Design:	G, TCPA

C	D	E	F	G
IS1	Design1	Relation	IS2	Design2
ESH	*	<	AH, BH, CH	*
Kol	G, TCPA	>	*	G, TCPA
LF1, U	*X	<	*	*X
ESP	*	>	ESH	*

Fig. 4. Transcription of the Expectancy relation to an Excel sheet.

Different to CogTool, that inherently ends up with similar HMI models fed into the simulation, map-centric monitoring HMI models for the same design can substantially differ between users and therefore the resulting predictions as well.

Therefore, we re-implemented the entire backend of CogTool to use our cognitive architecture CASCaS (Cognitive Architecture for Safety Critical Task Simulation) [9], [17, 18] and integrated support for generating operator models. These models support the Adaptive Information Expectancy (AIE) model [5] to predict the distribution of attention and the average reaction time to a certain event. Further on, we exchanged the hard-coded widgets palette of CogTool to annotate the designs with a model-based backend that enables us to define new annotation options (like for the IS in this case) without recompiling the tool.

The HEE currently supports users with the IS identification, automates the operator model generation and the generation of visualizations of the predicted attention distribution and the computed average reaction times (cf. Fig. 1).

Fig. 3 shows the HMI design modeling perspective of the HEE. The left vertical bar contains a palette of widgets that can be used to annotate a design. For attention allocation prediction, one widget ("i") is used to mark relevant sources of information for each design. The main area of Fig. 3 displays an already annotated design where several IS have been already marked by the user. Further widgets are used from the palette to identify the operator's initial line of sight, the operator's distance to the monitor, and to set the physical dimensions of the monitor.

For specifying the expectancy and value coefficients we use paper forms. We realized that those give the user more freedom by using quantors (c.f. subsection 3.1) to

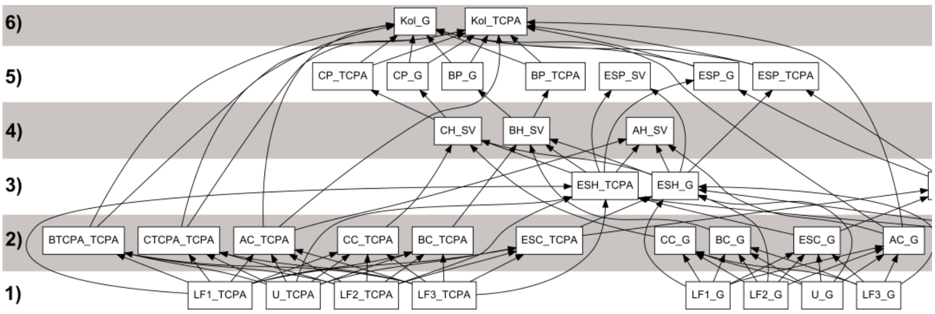


Fig. 5. Section of a Hasse diagram depicting the partial ordered graph with ranks on the left side. Node labels contain the abbreviation of an IS and a design name, separated by underscore.

aggregating relations over several IS and designs or by introducing abbreviations for IS that the user feels comfortable with. The expectancy is specified by relations like described in subsection 3.1. The value is defined by filling in a printed relevance matrix form (c.f. subsection 3.2). Both are then transcribed to an interactive excel sheet (see Fig. 4). To ease writing down the relations, we used abbreviations for the IS names, e.g., 'AH' for 'Vessel A Heading'. The sheet is then processed by a script that solves the quantors (c.f. subsection 3.1) and generates two Hasse diagrams: one for the expectancy (c.f. Fig. 5) and another one for the value coefficient. The graph ranks (c.f. Fig. 5) correspond to the coefficient values and then are entered in the HEE tool. After the user has chosen an IS to fire an event at a certain moment in time (c.f. subsection 3.3) the operator model for CASCaS can be automatically generated. It consists of:

- An HMI model defining the physical dimensions and spatial locations of the simulated operator and all IS (which we call topology in CASCaS). The HMI model is automatically created based on the rectangular definitions of information sources. The 2-dimensional coordinates are projected onto a 3-dimensonial plane, which is placed at a specific distance in front of the head of the simulated operator.
- An operator model composed of the operator's monitoring goals as a procedural knowledge specified as rules (c.f. subsection 3.4).

Both models are fed into CASCaS and executed as a Monte-Carlo simulation. Based on the desired visualization the HEE processes the simulation results and generates a heat map for each design (see Fig. 7). It is also possible to feed the data into interactive excel sheets that currently can generate a diagram with the average attention allocation per IS in percent as a pie chart and display the reaction time to an event while monitoring the interface as a histogram (see Fig. 8).

5 Use Case

Electronic Chart Display and Information Systems depicting nautical maps enriched by navigation-relevant information are one of the main sources of information of shipmasters and navigation officers and need to be continuously monitored.

Within the Cooperative Shipping and Navigation on Sea project, COSINUS[1], several new display designs have been proposed. Some of them are depicted in Fig. 6. Fig. 6 a) shows a state of the art chart display and Fig. 6 b) the corresponding design sketch, that highlights speed vectors (**SV**) as the current state of the art concept to display vessel behaviors with arrows and a red line illustrating the route of the own ship. All sketches illustrate a prediction. Therefore the speed vectors point to locations of the future vessels' locations in 10 minutes (with the assumption that they do not change their course and speed).

Fig. 6 c) depicts the design **G**: the entire routes of all surroundings vessels together with the maximum tolerance for deviations (gates), which have been set beforehand.

[1] http://www.emaritime.de/projects/cosinus, last checked 04/15/15

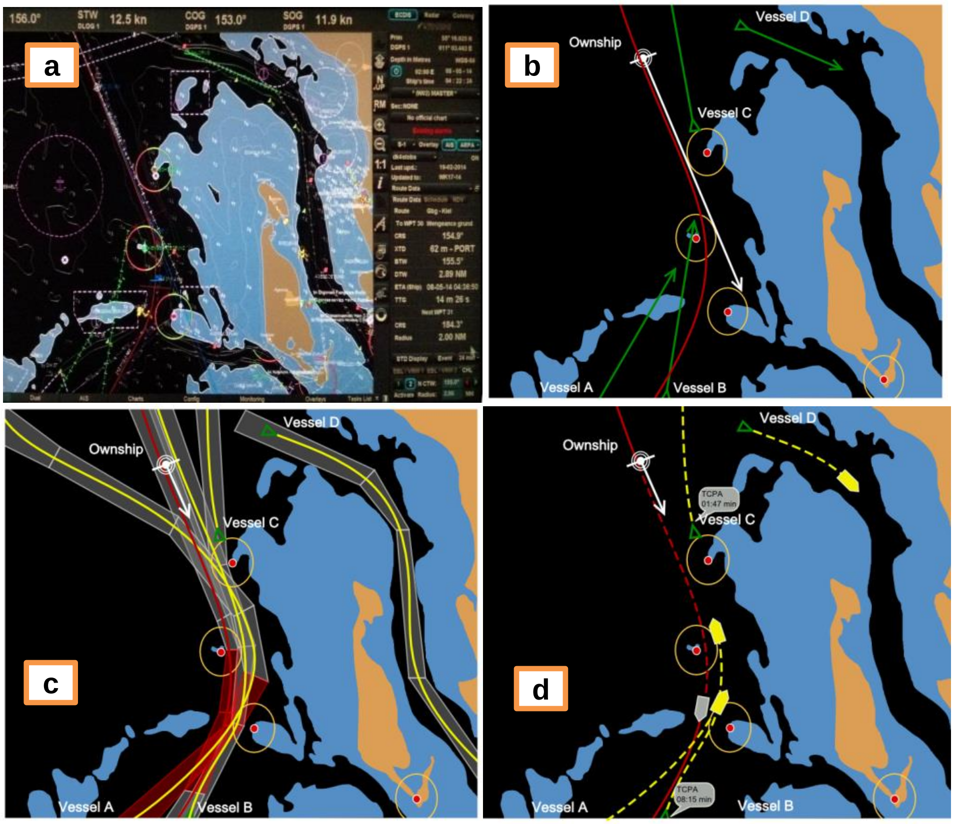


Fig. 6. Different variants of showing future vessel positions on a map display: (a) Current version, (b) Abstracted design of the map display: Speed Vectors, (c) Gates, (d) TCPA and position prediction.

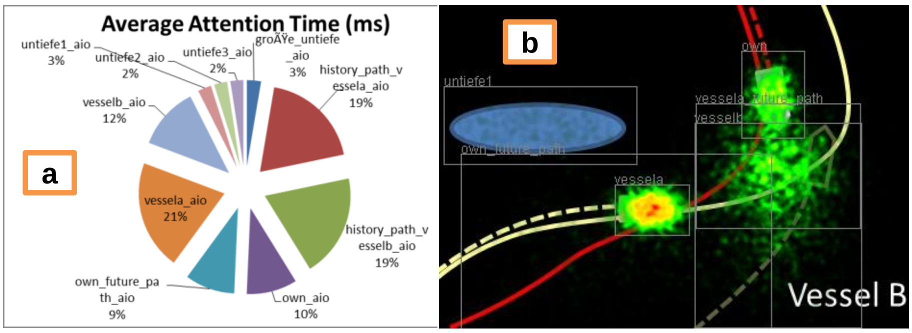


Fig. 7. Pie chart of the average attention allocation (a) and a heat map illustrating the monitoring behavior of the simulated operator (b).

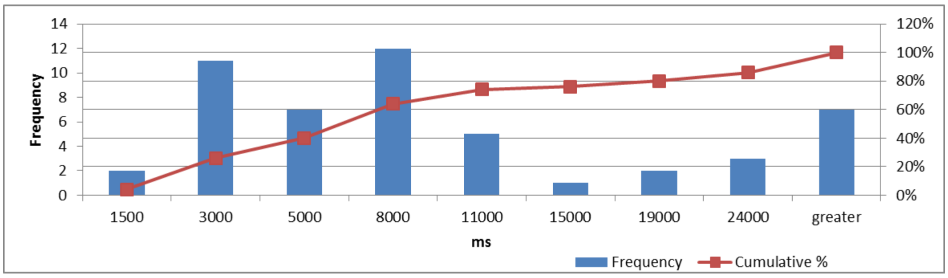


Fig. 8. Histogram of the Average Reaction Time. With a probability of 80% the average reaction time of the operator to a certain event took between 0 and 19s.

Regions where the temporal and spatial distance to other vessels are very small are marked red. The design **TCPA** of Fig. 6 d) does not present the entire route, but only the future routes of the surrounding vessels. These routes end with displaying the vessels' orientation and position in 10 minutes. Furthermore the time to closest point of approach (TCPA) is annotated at the vessels position.

6 Evaluation

We evaluated the attention prediction process with the HEE by a qualitative study[2]. The main objective was to get first insights into how well the HEE tool can be operated by users, and how helpful the visualizations that HEE creates are for the users. However, this is the first evaluation made for HEE. It therefore has a strong explorative character. Based on its findings, we plan to perform more extensive studies afterwards. The current study had a fixed, scripted procedure and implemented a within-subject design with four subject matter experts. The procedure contained explorative questions and pre-scripted help text (elaborated during a pre-test) that were planned to be read in case a subject is not able to perform a task. We designed the study to test three main hypotheses that were supported by several sub-hypotheses:

H_1:"Users without specific prior knowledge are able to use the HEE and end up with results in a reasonable amount of time."

H_2:"The variations between the models specified by the participants are small."

H_3:"The result visualization of the HEE is clear."

- **H_{3a}:**"A pie chart is an easy understandable visualization of the average attention allocation prediction."
- **H_{3b}:**"A histogram is an easy to understand visualization of reaction time prediction."
- **H_{3c}:**"Heatmaps are an easy understandable visualization of the average attention distribution prediction."

---

2 A video documentation of the study can be found online at http://multi-access.de/512

6.1 Participants

Four subject matter experts participated in the study: a cognitive modeling expert (**Cog**), an interface designer (**HMI**) that created the TCPA design sketch, and an expert for the maritime domain (**Exp**), and an expert for situation awareness (**SA**). All were between 30 and 55 years and neither had prior experience with using the HEE nor performed any of the three modeling tasks of the study before.

6.2 Apparatus

Participants were sat comfortably in front of a computer screen, and the experimenter sat to the side. A current state of the art map display, the three design variants (c.f. Fig. 6), and three basic tasks of a ship master (avoid low water sections; avoid dangerous vessel approaches; follow your planned route) were presented as slides to the subjects. The explanation was read from a manuscript.

6.3 Software

The HEE was opened with a project were the three designs variants have been already added. The subjects received an explanation of the modeling functions: marking an IS, resizing, moving, deleting, and naming it.

6.4 Procedure

The experiment was composed of the four expert activities of the overall prediction process (c.f. Fig. 1): (1) IS annotation with the HEE, (2) specifying the expected information update frequency, (3) task ranking by impact and defining the relevancy of each IS for each task and design variant, (4) identifying the IS of a predefined event. Further on, (5) the three resulting visualizations were analyzed. None of the participants was informed about what kind of results the HEE produces.

Between the first four activities that took each participant between 120-180 minutes in total and the fifth activity that took around 30-60 minutes there was a break. The break was required by the experimenters to calculate and prepare the visualizations.

6.5 Results

Based on the video recordings and notes taken during the study we evaluated the data with respect to each sub-hypothesis.

Table 1. Amount of time required to explain and to perform each modelling task in minutes.

Subject	IS Identification		Expectancy		Relevance		Entire Modelling Time	Entire study
Cog	3	23	6	16	1	3	42	94
SA	6	70	3	55	1	7	132	112
HMI	5	52	2	53	1	7	112	105
Exp	6	90	3	54	1	16	160	231

H_1:"*Users without specific prior knowledge are able to use the HEE and end up with results in a reasonable amount of time.*"

Table 1 shows the overall duration of the entire study for each participant and the total time for performing the three modeling steps. None of the subjects had used the tool before. Thus, for each step we gave a scripted introduction and allowed questions before each modeling activity started. The instruction time is listed in the first column of each modeling step. The mean of the modeling time is 2:02 hours. The Cog felt most familiar with all modeling steps and intuitively decided to mark the fewest IS (18) of all participants which reduced the work in the proceeding steps substantially. The SA had the most IS marked (47). We all asked the participants to comment on each IS identified, which Exp did in detail. It explains the relative high amount of time for the IS identification step (90min) of Exp.

$H_{2:}$"*The variations between the models specified by the participants are small.*"

As a first step we compared the IS defined by participants. All participants together defined 130 IS over all designs. During phase 1 participants explained which information they marked with an IS. We created a list of all this information and checked for each participant if it was marked. An excerpt of this list is shown in Table 2. The list contains 68 information elements. Sometimes the classification is not clear, because one participant marked an IS containing several information elements, while another participant marked one IS for each of these information elements. Therefore we distinguished between marking identical information, marking similar information and not marking the information.

Table 2. Comparison of IS between participants (excerpt).

Information	Cog		SA		HMI		Exp	
	identical	similar	identical	similar	identical	similar	identical	similar
Position beacon center-south	LF1		L1		LF3		L4	
Position beacon south-east			L4				L3	
Ownship Position	ESP		E		POwn		E	
Vessel A Position						RPLAB	A	
Vessel B Position	BP		V3			RPLAB	B	

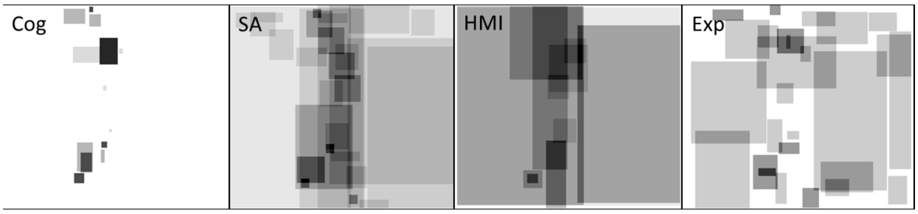


Fig. 9. IS marked by the subjects for design G. Darkness of the regions corresponds to the calculated expectancy coefficients.

Only four information elements are marked identical by all participants: 3 out of 4 beacons, the position of ownship and an area with higher danger of collision around the strait in design TCPA. Furthermore 8 information elements are marked by all participants, but not in the same way. These are headings and routes of vessels A, B and C and positions of vessels B and C. In contrast, 26 IS are only marked by one of the participants. Thus the IS models show a high variance.

There are not only differences in the number and kind of information that is marked. Even if the same information is marked by two participants, the marked regions of their IS can be very different. Fig. 9 shows for all four subjects the marked IS on design G. It is easy to visually identify differences between the four models. For example it can be seen, that the expert user marked only few small regions in the center of the sketch, while the HMI designer has covered the entire sketch with a focus on the center. As a measure of the similarity between rectangular IS we use the root integrated squared distance (RISD), which is sensitive to differences in size and position of two rectangles. Based on Table 2 the mean RISD between two IS from different participants marking identical information is far smaller ($\bar{M}$=66.8 mm), than for two IS marking similar information ($\bar{M}$=208.2 mm). For every information in Table 2 that was marked at least by two participants, we calculated the mean RISD between the IS. Not surprisingly it turned out, that the 10 most similarly marked information elements (mean RISD ≤ 28.7 mm) are displayed as small icons with clear boundaries: vessel positions, beacon positions, TCPA speech bubbles and labels. However, the 11^{th}-most similar information is an area with high danger of collision in design TCPA. This is surprising, because there is no clear symbol or display on the screen that gives a boundary for this area. All participants marked it solely based on interpretation and aggregation of several other information elements, like predicted vessel positions, routes and strait. It is also one of the very few information elements, which was marked by all four participants. It seems that the TCPA design easily supports the recognition of critical regions, without directly highlighting them.

As a conclusion, we did not find support for the hypothesis, because the number of defined IS greatly varies and also the kind of marked information elements. Furthermore, even the same information was marked differently. However, this difference varies strongly between the information elements.

H$_3$: *"The result visualization of the HEE is clear."*

H_{3a}: *"A pie chart is an easy understandable visualization of the average attention allocation prediction."*

Confronted with the average attention allocation prediction, visualized as a pie chart (Fig. 7a) and without knowing which pie chart reflects which design, all subjects raised the question of how an optimal attention distribution can be identified. Three subjects assumed that most attention should be focused on a few IS only, only HMI argued for a "balanced distribution". Cog assumes "reduced scanning effort" for few big chunks although mentioned that huge chunks might identify IS which require high effort to derive specific information. SA stated that huge chunks combined with only a low amount of IS identify less complex information spaces, but in this case it remains questionable if a small information space still offers all information to perform a task. Different to the others, only Exp analyzed the specific IS that received most attention. Exp considered the own vessel direction and speed as most important IS and preferred to have the attention focused on them. Based on observing the four subject matter experts (SMEs) the pure pie chart did not support H_{3a}. The biggest problem was the lack of measure to judge about the goodness of overall attention allocation.

H_{3b}: *"A histogram is a simple to understand visualization of a reaction time prediction."*

Although each participant was given an example on how the cumulative distribution function is read (e.g. "In 80% of the simulated cases the reaction time was below 10 seconds"), they focused on the frequency distribution shown as bar diagram, whereas Cog required additional support while initially focusing on the cumulative frequency distribution (line chart). Based on the four subjects we could not found support for H_{3b}.

H_{3c}: *"Heatmaps are an easy understandable visualization of the average attention distribution prediction."*

All participants stated that the heat map-based visualization of the attention distribution matches their expectancies. Only for design G HMI missed a hot spot for the crossing gates, Cog missed attention for one vessel and SA could not explain strong focus on the own ship in one design. While looking at the heat map all participants found arguments for their preferred design. The decision for one design based on the heat map visualization was difficult for three participants. SA decided based on the reduction in ergonomic effort (spots are located close to each other) and also looked for equally distributed hotspots of the most relevant IS. HMI based the decision on looking for hotspots covering the most relevant IS. Cog and Exp took their decision instantly based on the design that hot spotted the high traffic area best while Exp also considered the reliability of the IS. We could found support for H_{3c} since all SMEs understood the visualization, stated that it presents expected hotspots and were able to argue based on the visualization.

7 Discussion

In earlier studies we used the HEE to predict task performance and operator workload for HMI designs in the aeronautics domain and the Adaptive Information Expectancy (AIE) model to predict attention distribution in the automotive domain. The analysis of map-centric monitoring tasks is based on different prerequisites. The model creation figured out to be much more based on individual experiences and background unlike for other HMIs: In an automotive HMI or an airplane cockpit instruments communicate clearly defined information and each instrument has a fixed position. Monitoring a nautical map involves interpretation and the amount, size and position of ISs considered as relevant differ greatly between the roles that participate in the design process.

This has an effect on the attention predictions. The SMEs modeled their perspective and understanding of the HMI quite differently and with different inputs the attention predictions results differ as well. Interestingly, this does seem to affect the users' expectancy: None of our participants was really surprised by the results. Especially for the heat maps it could be observed that the predictions offered only few inconsistencies from what was expected and all subjects were able to analyze and argue about the designs (H_{3c}) based on their results. Furthermore the overlay of IS on the design sketches were used by participants during arguing to point at exactly the information they were talking about.

We summarize our observations into the following hypotheses:

Hp:"*Using the HEE for modeling map-based monitoring tasks makes expert knowledge explicit and can be used as a basis for argumentation of a role specific perspective."*

Hp_a:"The resulting models differ between different roles."
Hp_b:"Role-specific IS can be identified and visually communicated."
Hp_c:"The definition and naming of IS provides a common vocabulary, which supports the discussion between different perspectives."

The average reaction time visualization as a histogram figured out to be complicated to understand (H_{3b}). It was hard for the subjects to identify the shortest reaction time. The probability and their respective cumulation were difficult to understand for the subjects. The visualization of the attention allocation as a pie chart did not motivate the subjects to reflect the percentage share of specific IS of the overall attention allocation. Rather the subjects were unsure about a concrete measurement to argue for or against big chunks or for comparing slide sizes (H_{3a}).

8 Conclusion and Future Work

In this paper we presented a tool-supported, semi-automated process to predict human attention allocation that enables non-experts in cognitive modeling to analyze nautical chart displays. In a qualitative study experts with different backgrounds were able to understand and successfully perform the process and generate HMI-models of three different HMI design variants in a reasonable amount of time.

We initially assumed that the HEE tool helps non-human factor experts to create valid predictions of human attention distribution to map-based HMIs. In contrast this study revealed that different HEE users created very different models, resulting in different predictions. The subjects identified a total amount of 130 information sources, only 4 are equal, 8 more are similar but marked differently. 26 were only identified by one of the participants.

We also tested different types of result visualizations. The visualization of the predicted reaction time distribution by a histogram figured out to be too complicated to understand. Similar, pie-charts depicting the average attention allocation offered only little use for most of the subjects. However, heat maps of attention distributions were understood by all subjects. It turned out that information depicted in the heat maps was in the range of what was expected or learnt by the participants during modeling. We further observed that HEE users easily articulate their respective expert knowledge while referencing their marked IS on the heat map.

It is technically easy to identify differences between models created by different users. We assume that showing these differences with a suitable visualization can help to exchange knowledge between experts with different background and also differences in their mental models can be revealed. Furthermore the explicit definition of IS could help experts of different areas to communicate with a shared vocabulary. These assumptions will be investigated in future studies.

Acknowledgments. The HEE is developed in the EU ARTEMIS JU project HoliDes (http://www.holides.eu/) SP-8, GA No.: 332933. Any contents herein reflect only the authors' views. The ARTEMIS JU is not liable for any use that may be made of the information contained herein.

9 References

1. International Maritime Organization. Annex 24 – MCA guidance notes for voyage planning. Technical report, IMO RESOLUTION A.893(21), https://mcanet.mcga.gov.uk /public/c4/solas/solas_v/Annexes/Annex24.htm, 1999, last checked 04/15/15.
2. Michelle R Grech, Tim Horberry, and Andrew Smith. Human error in maritime operations: Analyses of accident reports using the leximancer tool. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 46, pages 1718–1721. SAGE Publications, 2002.
3. Nathaniel Bowditch. *The American practical navigator : An Epitome of navigation*. Paradise Cay Publications, 2002.
4. Erik Hollnagel and David D. Woods. *Joint cognitive systems : foundations of cognitive systems engineering*. Taylor & Francis, Boca Raton, FL, 2005.
5. Bertram Wortelen, Martin Baumann, and Andreas Lüdtke. Dynamic simulation and prediction of drivers' attention distribution. *Transportation Research Part F: Traffic Psychology and Behaviour*, 21:278–294, 10 2013.

6. Christopher D. Wickens, Juliana Goh, John Helleberg, William J. Horrey, and Donald A. Talleur. Attentional models of multitask pilot performance using advanced display technology. *Human Factors*, 45(3):360–380, 2003.
7. Bertram Wortelen. *Das Adaptive-Information-Expectancy-Modell zur Aufmerksamkeitssimulation eines kognitiven Fahrermodells*. PhD thesis, Universität Oldenburg, Fakultät II Informatik, Wirtschafts- und Rechtswissenschaften, Department für Informatik, 2014.
8. Bertram Wortelen and Andreas Lüdtke. Adaptive simulation of monitoring behavior. In *The Sixth International Conference on Advances in Computer-Human Interaction*, 03 2013.
9. Andreas Lüdtke; Jan-Patrick Osterloh; Florian Frische. Multi–criteria evaluation of aircraft cockpit systems by model–based simulation of pilot performance. In *Proceedings of ERTS - Embedded Real Time Software and Systems 2012*, 02 2012.
10. Christopher D. Wickens, Jason S. McCarley, Amy L. Alexander, Lisa C. Thomas, Michael Ambinder, Sam Zheng. Attention-Situation Awareness A–SA Model of Pilot Error. In David C. Foyle, Becky L. Hooey, editor, *Human Performance Modeling in Aviation*, pages 213-239. CRC Press, New York, 2008.
11. Chris Forsythe, Michael L. Bernard, and Timothy E. Goldsmith, editors. *Cognitive Systems: Human Cognitive Models in Systems Design*. Psychology Press, 2006.
12. Christopher Wickens, Angelia Sebok, John Keller, Steve Peters, Ronald Small, Shaun Hutchins, Liana Algarín, Brian F. Gore, Becky L. Hooey, and David C. Foyle. Modeling and evaluating pilot performance in nextgen: Review of and recommendations regarding pilot modeling efforts, architectures, and validation studies. Technical report, Human Centered Systems Lab, 2013.
13. Bonnie E. John, Konstantine Prevas, Dario D. Salvucci, and Ken Koedinger. Predictive human performance modeling made easy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 455–462, New York, NY, USA, 2004. ACM.
14. John R. Anderson and Christian J. Lebiere, editors. *The Atomic Components of Thought*. Lawrence Erlbaum Associates, Mahwah, NJ, 1998.
15. Kevin M. Corker and Barry R. Smith. An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation automation. In *Proceedings of AIAA Computing in Aerospace 9 Conference*, San Diego, CA, 21 October 1993.
16. Christopher D. Wickens, John Helleberg, Juliana Goh, Xidong Xu, and William J. Horrey. Pilot task management: Testing an attentional expected value model of visual scanning. Technical report, NASA Ames Research Center Moffett Field, CA, 2001.
17. Andreas Lüdtke, Lars Weber, Jan-Patrick Osterloh, and Bertram Wortelen. Modeling pilot and driver behavior for human error simulation. In Vincent G. Duffy, editor, *Digital Human Modeling*, volume 5620 of *Lecture Notes in Computer Science*, pages 403–412. Springer Berlin Heidelberg, 2009.
18. F. Frische, J.-P. Osterloh, and A. Lüdtke. Simulating visual attention allocation of pilots in an advanced cockpit environment. In *Modelling and Simulation (MODSIM) World Conference & Expo*, Hampton, Virginia, USA, 13– 15 October 2010.