Modeling Monitoring Behavior for HMI Designs is Easy with the Right Tool

Bertram Wortelen[1], Sebastian Feuerstack[2], Marcus Behrendt[2]
[1]Carl von Ossietzky University, 26129 Oldenburg
bertram.wortelen1@uni-oldenburg.de
[2]OFFIS – Institute for Information Technology, 26121 Oldenburg, Germany
feuerstack@offis.de, marcus.behrendt.86@gmail.com

Abstract. Monitoring behavior is usually measured by performing studies in that eye-movements of operators monitoring a real system are recorded. Modern eye-tracking systems generate heat maps from recorded data. Areas of interest and the corresponding attention distribution can then be derived from the heat maps. We argue that recent progress in research indicates that psychological and physiological plausible human behavior models can often substitute time-consuming eye-tracking studies. Further on, with prediction models and the right tooling, heat maps can already be generated in early design phases: based on design sketches instead of running prototypes. We present a case study in the maritime domain in that experts predicted and analyzed their monitoring behavior. We argue that with the right tooling even non-experts will be able to predict their monitoring behavior. Enabling easy access to monitoring behavior prediction for everyone will improve future HCI design.

1 Introduction

Supervision and monitoring of complex systems is one of the main activities of an operator in a safety-critical environment such as air traffic control rooms, air plane cockpits, or ship bridges. More and more systems like cars, our office and home environments are getting smarter and act autonomously on behalf of us. This changes our role from being in active control to being there to observe and understand what is going on.

This has an impact on interface design. Being aware of how a design change might affect the human monitoring behavior is important information. Today the human monitoring behavior is often analyzed by recording eye movements, which requires observing humans controlling a working prototype of an interface design. We argue that psychological and physiological plausible human behavior models can often substitute time-consuming eye-tracking studies, can be earlier performed based on design sketches, and, if supported by a tool can also be performed by non-experts. In fact, analyzing and predicting human monitoring behavior has been a research topic for several decades. In the mid of the 20th century a series of studies were conducted to analyze monitoring behavior of fighter jet pilots [3] in order to design cockpits with an optimized layout.

Since then, several models have been developed to predict human monitoring behavior in workplace environments like control rooms, cockpits, etc. In these environments the human behavior is strongly shaped by the tasks of the operator. A common prediction model assumption is, that humans try to perform tasks in an optimal way [6, 2, 8, 7, 5, 10, 1, 13]. Nearly all of these models argue that optimal monitoring behavior is based on knowledge of the probability distribution of information events for each information source and knowledge about the value of perceiving information events or respectively the costs of missing events. Wickens et al. [10] describe this as the two knowledge-based forces expectancy and value (of information events) that affect attention distribution. Besides these forces there are extrinsic factors like salient information events and the effort associated with sampling an information source that can lead to a potentially negative deviation from optimal monitoring [10, 9]. However, with sufficient training the knowledge-based factors should explain the majority of variance in monitoring behavior [5, 11].

Even though there is a vast knowledge base on monitoring theory, the above mentioned models are typically only used by human factors experts, because in order to use any of the above mentioned models to predict optimal monitoring behavior, a quantification of expectancy and value coefficients for each information source is required. We believe that this is a barrier for a more wide-spread adoption of attention models in industrial practice and we believe that a suitable software tool can enable non-experts to do this task. Furthermore we believe that even if the modeling predictions are not very accurate the modeling process itself is useful for explicitly extracting the knowledge of experts in a structured way.

In this contribution we present a tool for the quantification of value and expectancy coefficients for a given set of information sources. These coefficients can directly be used for the value and expectancy factors of Wickens' SEEV [10] and A–SA [11] models or the recent implementations of the SEEV model in the cognitive architectures CASCaS [12] and MIDAS [4]. The next section describes a case study for monitoring behavior prediction in the maritime domain that was partially tool-supported. Based on the experiences gained, Section 3 proposes a complete tool-supported prediction process to facilitate the quantification of coefficients that is targeted also to non-experts. Section 4 summarizes our position.

2 Predicting Maritime Chart Monitoring Behavior

Three different design variations for a nautical chart display were evaluated to test the effect that offering a new kind of information has on operator's monitoring. Four experts with different backgrounds participated in this qualitative study [3]. All were able to understand and successfully perform the process of generat-

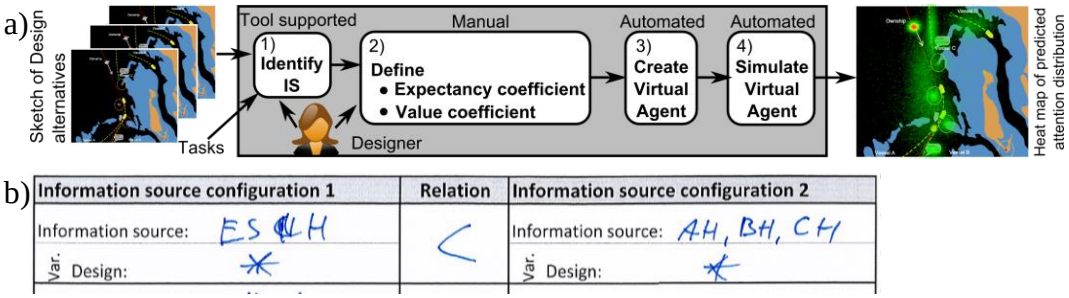

Fig. 1. (a) Tool-supported process to predict operator's attention allocation. It was executed by the HMI designer and the operator [3]. (b) Paper form to specify expectancy relation.

ing attention predictions. Fig. 1 illustrates the steps that the participants performed. In the first step, which already was tool-supported, they visually marked all sources of information (IS) on each design sketch. Different to what we expected, a high amount of IS were identified (between 18 and 47) and it figured out hard for them to remember all of them in the subsequent steps. We therefore supported them by a list with all IS identified. We also introduced abbreviations for each IS.

The second step (expectancy and value identification of all IS) was performed manually by filling paper forms (c.f. Fig. 1b). The expectancy of information events occurring in an IS was identified by mathematical relations ("<": greater, ">": smaller, and "=": identical) between IS that are relevant for a predefined set of tasks. For example the relation depicted in Fig. 1b states that IS "ESH" is expected to provide less new information than IS "AH", "BH", or "CH" (for all design variants). Since the same IS sometimes occurred on more than one design, with each IS also the corresponding design name was stated in the relation and we introduced a "*" operator, which could be used to refer to an IS on all designs. In a similar way a "*" operator can also be used to refer to all ISs on a specific design.

Unfortunately due to high number of ISs and complexity of the relation, one participant created an inconsistent relation that we were not aware of during the study. We could observe that all participants had trouble with the "mathematical" way of specifying the relation: It took them time to get comfortable with the notation and some required help to correctly write down what they intend to say. After a participant was finished with defining the relation, we systematically walked though all IS identified and asked if this IS has or needs to be considered in the relation. This was done to ensure that the participant has at least considered all IS as candidates for including them in the relation. In a similar way, participants rated operator tasks by creating a task relation based on the task value. Additionally the relevance of an IS for a certain task was rated as either "necessary", "helpful", or "not relevant" on a form. All forms were then transcribed into an

Excel table. Expectancy and value coefficients were calculated using the lowest ordinal algorithm [6, 15, 13], which derives the coefficients from the rank order of the IS in the expectancy relation, respectively the task in the task value relation.

The predictions were generated in a reasonable amount of time (between 42 and 160 minutes modeling time, without transcription time). Between 44% and 54% of this time was spent on identifying the expectancy relations, values of each IS, and by filling the paper forms. The next section details the tool-support and states how we considered the study observations for implementing a complete tool-supported attention prediction process.

3 Towards a structured, tool-driven process to compute attention predictions

We think that a complete tool-driven attention prediction process can resolve all issues that required our intervention in the maritime study, will reduce the overall modeling time, and also will enable non-experts to create attention predictions. The IS identification step was already tool supported in the study, thus we will focus on the tool-based generation of the expectancy and value coefficients. Based on the time measurements of the study the former one required most parts of the modeling time for all participants, whereas the value definition only took 8.2 minutes on average [3].

We decided for a two-step approach to define the expectancy relations: First, the user is asked to roughly build up a hierarchy of IS by dragging them from a list (Fig. 2c) into a hierarchy (Fig. 2b). Relations are automatically created while the hierarchy is established (Fig. 2d - for each two consecutive layers in the hierarchy a relation statement is created). In the middle of each relation statement there is a button indicating the current operator that can be clicked to toggle between “<”, ”>”, and ”=”. Second, new statements can be added and existing ones can be updated by directly editing the relation list of relation statements (Fig. 2d). The list of all IS (which is the source for dragging out “IS”) can be sorted alphanumerical, per design (like shown in Fig. 2c) and most importantly can be sorted by “usage” to ease identifying IS that have not been defined in a relation statement so far. Inconsistent relation statements are instantly highlighted red (e.g. the last one in Fig. 2d).

We observed during the study that the participants intuitively grouped similar IS logically to reduce the writing effort (e.g. all “lighthouses” or all “shoals” in the study). Therefore we introduced IS grouping support to the tooling: The selection and dragging of several IS from the IS list (Fig. 2c) ends up in a popup

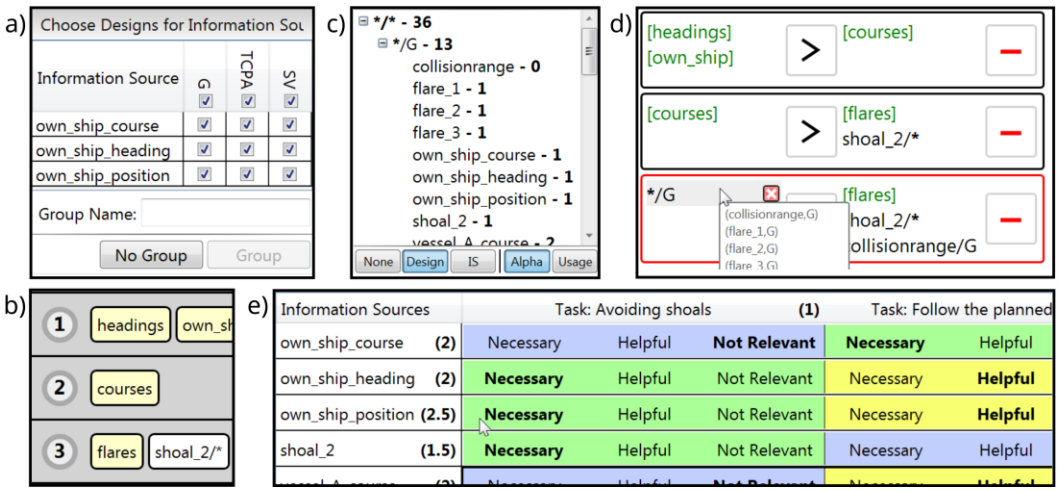

Fig. 2. Tool screenshots. (a) Popup window for defining an IS group. (b) Hierarchy of ISs. (c) List of all ISs. (d) Three expressions of an expectancvy relation. (e) Relevance matrix.

window (Fig. 2a) enabling a fine grained selection of the designs (columns) that should be part of a group. Additional, a group name can be defined (e.g. "own_ship" in Fig. 2a) that is then used in the relations (Fig. 2d – groups are identified by square brackets). Having the mouse pointer hovering above a group name or any other shortcut operator (like the "*" operator in Fig. 2d that sums up all IS from Design "G"), a tooltip shows all IS inside the group.

For the definition of the IS, we kept the idea of the paper form that used a relevance matrix to rate the value of each IS for a certain task with either "Necessary", "Helpful", or "Not Relevant" (Fig. 2e). Since the study participants were already fast with the rating process, we only added a coloring scheme to ease the identification of IS that have not been rated and also to support the user in perceiving, which tasks require a lot of IS or which IS is relevant for many tasks.

After the relations and relevance matrix have been defined the consistency of relations and matrix are automatically checked. Then the expectancy and value coefficients are calculated and fed into the following process step for creating the virtual agent. Manual transcription is no longer required. Up to this point the user worked solely with the mouse, except for naming the groups. The need for writing down the relation in the mathematical notation is avoided. Initial relations are automatically created when setting up the expectancy hierarchy (Fig. 2b). We expect that this eases the understanding of the basic structure of the expectancy relation. The manual addition of new relations is also based on drag-and-drop operations between the IS list (or the hierarchy view) and the fields of the relations to further reduce typing effort.

4 Conclusion

We presented a tool that supports all steps required to characterize optimal monitoring behavior. All actions that the users have to do are kept as simple as possible. In fact, besides the domain knowledge of the monitoring task we expect that very little knowledge about the underlying monitoring model is required. The calculation of expectancy and value coefficients is automatically done by the tool. The required domain knowledge is extracted by guiding the user through a series of actions. Each statement of the relation (c.f. Fig. 2d) or entry in the relevance matrix captures a single aspect of the domain knowledge, e.g., "information events occur more frequently in information source A than in information source B" or "information source C is helpful for performing task D". The integration of all these statements in an overall model and the calculation of the resulting coefficients is entirely performed within the tool.

In a follow-up action, we will conduct a study in the automotive domain to test how valid the predictions of the users are. We will ask experienced drivers to characterize monitoring behavior in a set of different driving situation using our tool. We intend to only offer a short video tutorial to teach the participants the tool usage. In parallel a driving simulator study is conducted for all of these driving situations. The predicted attention distribution will be compared with drivers' attention distribution measured with an eye tracker. Other studies [6, 15, 13] that used human factor experts and the lowest ordinal heuristic to calculate the value and expectancy coefficients reported correlation coefficients for percentage dwell times for all information sources in the range of $0.60 <= r <= 0.98$. These studies were performed on a similar level of abstraction in realistic driving and flight simulators. On average over all participants we expect to achieve a correlation within this range.

References

[1] Byrne, M.D., Kirlik, A.. Integrated modeling of cognition and the information environment: A multilevel investigation (process and product modeling) of attention allocation to visual displays. Technical Report AHFD-04-14/NASA-04-4, Aviation Human Factors Division (AHFD), Institute of Aviation, University of Illinois, Urbana-Champaign, Savoy, Illinois 61874, Sept. 2004.

[2] Carbonell, J.R.: A queuing model of many-instrument visual sampling. IEEE Transactions on Human Factors in Electronics, 7(4):157–164, Dec. 1966.

[3] Feuerstack, S., Wortelen, B.: Revealing differences in de-signers' and users' perspectives: A tool-supported process for visual attention prediction for designing HMIs for maritime monitoring tasks. In Proceedings of 15th IFIP TC.13 International Conference on Human-Computer Interaction - INTERACT, Bamberg, Germany, 14.-18. Sept. 2015.
[4] Fitts, P.M., Jones, R.E., Milton, J.L.: Eye movements of air-craft pilots during instrument-landing approaches. Aeronautical Engineering Review, 9(2):24–29, 1950.
[5] Gore, B.F., Hooey, B.L., Wickens, C.D., Scott-Nash, S.: A computational implementation of a human attention guiding mechanism in MIDAS v5. In Duffy, V.G. (eds) Digital Hu-man Modeling, vol. 5620/2009 of Lecture Notes in Computer Science, pp. 237–246. Springer, Berlin, 2009.
[6] Horrey, W.J., Wickens, C.D, Consalus, K.P.: Modeling drivers´ visual attention allocation while interacting with in-vehicle technologies. J. Exp. Psychol.-Appl., 12(2):67–78, 2006.
[7] Moray, N.: Monitoring behavior and supervisory control. In Boff, K.R., Kaufman, L., Thomas, J.P. (eds) Handbook of Perception and Human Performance. vol 2. Cognitive Processes and Performance, chapter 40, pp. 40–1 – 40–51. Wiley-Interscience, Nov.1986.
[8] Senders, J.W.: The human operator as a monitor and controller of multidegree of freedom systems. IEEE Transactions on Human Factors in Electronics, HFE-5(1):2–5, Sept. 1964.
[9] Sheridan, T.B.: On how often the supervisor should sample. IEEE Transactions on Systems Science and Cybernetics, pp. 140–145, 5 1970.
[10] Smallwood, R.D.: Internal models and the human instrument monitor. IEEE Transactions on Human Factors in Electronics, HFE-8(3):181–187, 1967.
[11] Steelman-Allen, K.S.: Modeling the Control of Visual Attention in Complex Workspaces. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois, 2011.
[12] Wickens, C.D., Helleberg, J., Goh, J., Xu, X., Horrey, W.J.: Pilot task management: Testing an attentional expected value model of visual scanning. Technical Report ARL-01-14/NASA-01-7, University of Illinois, Aviation Research Lab, Savoy, IL, Nov. 2001.
[13] Wickens, C.D, McCarley, J.S., Alexander, A.L., Thomas, L.C., Ambinder, M., Zheng, S.: Attention-situation awareness (A-SA) model of pilot error. In Foyle, D.C., Hooey, B.L. (eds) Human Performance Modeling in Aviation, pp. 213–239. CRC Press/Taylor & Fran-cis Group, New York, 2008.
[14] Wortelen, B.: Das Adaptive-Information-Expectancy-Modell zur Aufmerksamkeitssimulation eines kognitiven Fahrermodells. PhD thesis, University Oldenburg, 11 2014.
[15] Wortelen, B., Baumann, M., Lüdtke, A.: Dynamic simulation and prediction of drivers' attention distribution. Transportation Research Part F: Traffic Psychology and Behaviour, 21:278–294, 2013.