Effects of Task Performance and Task Complexity on the Validity of Computational Models of Attention

Lisette de Koning[1], Peter-Paul van Maanen[1,2], Kees van Dongen[1]

[1]TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
{lisette.dekoning, peter-paul.vanmaanen, kees.vandongen}@tno.nl
[2]Vrije Universiteit Amsterdam, De Boelelaan 1081a,
1081 HV Amsterdam, The Netherlands

Computational models of attention can be used as a component of decision support systems. For accurate support, a computational model of attention has to be valid and robust. The effects of task performance and task complexity on the validity of three different computational models of attention were investigated in an experiment. The gaze-based model uses gaze behavior to determine where the subject's attention is, the task-based model uses information about the task and the combined model uses both gaze behavior and task information. While performing a tactical compilation task, participants had to indicate to what set of objects their attention was allocated. The indications of the participants were compared with the estimations of the three models. The results show that overall, the estimation of the combined model was better than that of the other two models. Contrary to what was expected, the performance of the models was not different for good and bad performers and was not different for a simple and complex scenario. The difference in complexity and performance might not have been strong enough. Further research is needed to determine if improvement of the combined model is possible with additional features and if computational models of attention can effectively be used in decision support systems. This can be done using a similar validation methodology as presented in this paper.

INTRODUCTION

In the domain of naval warfare, information volumes for navigation, system monitoring and tactical tasks will increase while the complexity of the internal and external environment also increases (Grootjen & Neerincx, 2005). The trend of reduced manning is expected to continue as a result of economic pressures and humans will be responsible for more tasks with increased workload. Attention can be divided between different tasks, however attentional resources are limited (Wickens, 1984; Kahneman, 1973). Experience, training and interface design can improve these limitations, but only to a certain level. Even with experienced users, attentional problems are still likely to occur (Pavel, Wang, & Li, 2003). In naval warfare, errors caused by attentional problems can have serious consequences. Automation can assist the human by directing attention to critical events via alarms and alerts (Wickens, 2007) or by taking over tasks. Knowing when a user needs support, a cognitive model of attention can be used. A cognitive model of attention is a model that estimates where the attention of the user is allocated at a certain moment. Together with a normative model that estimates where attention should be allocated, a decision support system can aid the user in dividing limited attentional resources. When these models are not accurate, support occurs at the wrong place and the wrong time. This will affect trust and acceptance of the decision support system and subsequently reliance behavior (Parasuraman & Riley, 1997; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). To be able to develop a valid cognitive model of attention it is important to know what information can be used to estimate where attention is allocated and what factors affect the validity of a cognitive model of attention.

In the following section we will discuss what information is useful in estimating where attention is allocated. This will lead to three different cognitive models of attention. After that, we will discuss the effects of task complexity and performance on the validity of the three models and describe the corresponding hypotheses. The focus in this paper will be on the tactical picture compilation task that is performed in naval warfare. The goal of a tactical picture compilation task is to build up a situation of surrounding ships (contacts).

Allocation of attention

The direction of eye gaze is informative about where attention is directed (Just & Carpenter, 1976; Salvucci, 2000). In search tasks, eye movements may be indicative of where attention is allocated. A tactical picture compilation task is comparable with a search task. Targets (hostiles) have to be identified between different distractors (for example, container ships). In directing attention, a distinction has to be made between overtly orienting attention and covertly orienting attention. Overt changes in directing attention can be observed by head movements and eye movements. Covert orienting means directing attention to a location other than the one where the eyes are fixated and cannot be measured by eye and head movements (Posner, 1980). To be able to estimate where attention is allocated, gaze behavior will not be sufficient. Besides gaze behavior, knowledge about how a task is expected to be performed may also be indicative of where attention is allocated. The goal of a task will affect where attention will be allocated in a top-down manner (Treisman & Gelade, 1980) and will affect both covert and overt allocation of attention. When the goal of the task is to keep track of green objects, attention may be directed overtly to those objects, but when the targets are found it is possible to track those targets covertly. Information about the goal of the task will provided additional information about where attention is allocated next to eye-movements. Besides eye-movements and information about the task, saliency of different stimuli of the task (e.g., color or brightness) may also be informative about where attention is allocated. Salient objects may attract attention in a bottom-up manner (Ouerhani, von Wartburg, Hügli, Müri, 2004). However, it is expected that this effect will be minimal in a tactical picture compilation task. Features that might capture attention in a bottom-up fashion, for example high speed of a contact, are also expected to receive attention based on the goal of the task. Contacts with a high speed are more threatening than contacts with a lower speed and will therefore receive attention. To determine how informative gaze behavior, characteristics of the task and the combination of both are to estimate where attention is allocated, three different cognitive models of attention were developed. The first model, the gaze-based model, uses gaze behavior to estimate where attention is allocated. The second model, the task-based model, uses information about how the task is expected to be performed. The third model, the combined model, uses both types of information to estimate where attention is allocated. Task complexity and task performance affect allocation of attention and are also expected to affect the validity of the three cognitive models. In the following sections, the possible effects of these variables on the validity of the cognitive models will be discussed.

Task complexity

Task complexity is related to multiple features of a task, for example having to deal with rapidly evolving situations, cognitive complexity and uncertain data (Wood, 1986). Considering the characteristics of a picture compilation task, the complexity of a task can be determined by the *dynamics, ambiguity* and *volume* of information. Information is *dynamic* when the type, semantics, or volume of information varies over time. Information is *ambiguous* when the information which is needed to perform the task is unclear, incomplete, contradictive or inaccurate. The *volume* of information refers to the amount of information or events that occur at the same time. When task complexity increases, for example because of ambiguous information or more contacts, more attentional resources are needed to identify the contacts and it will be harder to allocate attention to the right contacts. Estimating where attention of the user is allocated is also more difficult. This leads to the following hypothesis: *The validity of all three models is higher in a simple than in a complex task (H1).*

The combined model uses more information to estimate where attention is allocated than the other two models, namely information about gaze behavior and information about how the task is expected to be performed. Using only one type of information will not be sufficient to estimate where attention is allocated. This leads to the following hypothesis: *For both complex and simple tasks, the validity of the combined model is higher than both the task- and the gaze-based models (H2).*

Use of more information is of extra value in complex tasks, because in complex tasks it is harder to estimate where attention is allocated. This results in the following hypothesis: *The difference in validity between the combined model and the task- and gaze-based model is higher in a complex than in a simple task (H3).*

Task performance

How well people perform a certain task affects the allocation of their attention. People that are more experienced will be better at dividing attention between different sources of information. Research on the effects of playing video games has shown that our visual attention abilities may improve with training. Experienced players of video games required less attentional resources for a given target (Green & Bavelier, 2003). When performing a tactical picture compilation task, experts will be able to track more contacts. Experts will also be able to determine more quickly whether a contact is a possible threat. As opposed to poor performers, good performers will apply the rules correctly. The allocation of attention of good performers will be very similar to the task-based model. For poor performers, the allocation of attention will differ from the estimate of the task-based model. The combined model is partly based on the task-based model. This results in the following hypothesis: *The validity of the combined and the task-based model is higher for good performers than for poor performers (H4).*

The combined model uses more information to estimate where attention is allocated than the other two models, namely information about gaze behavior and information about how the task should be performed. This leads to the following hypothesis: *For both good and poor performers, the validity of the combined model is higher than both the task- and the gaze-based models (H5).*

When hypothesis 2 is true and when hypothesis 5 is true, this leads to the following hypothesis*: The validity of the combined model is higher than both the task- and the gaze-based models (H6).*

METHOD

Participants

42 college students (22 male, 20 female) with an average age of 23 years (*SD* 2.29) participated in the experiment as paid volunteers.

Task

Participants had to perform a dual task. The first task is derived from the tactical picture compilation task that is performed in naval warfare. The goal of the tactical picture compilation task is to build up awareness of threats surrounding the ships (contacts). In Figure 1 a screenshot of the interface of the task environment is shown.

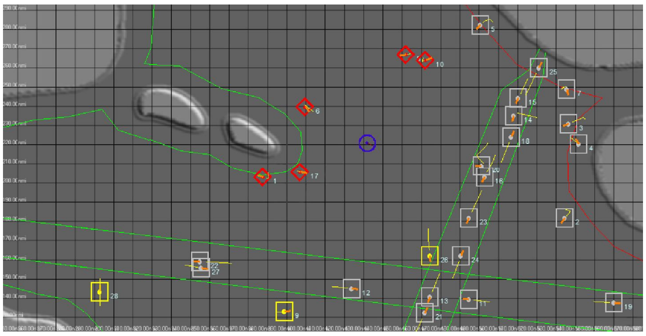

Figure 1. Radar screen with different contacts.

Participants had to identify the five most threatening contacts and mark them red by clicking on them. To determine if a contact is a possible threat the following criteria had to be used: speed, heading, distance of a contact to the own ship and whether a contact was positioned in a sea-lane or not. The behavior of the contacts was such that they varied on these criteria, which made them more or less threatening over time. For instance, a contact could get out of a sea lane, speedup, or change its heading toward the own ship. Contacts that were mistakenly identified as a threat (false alarm) or contacts that were mistakenly not identified as a threat (miss) resulted in a lower performance score. More details about the task environment are described in Van Maanen, de Koning, & van Dongen (2007). The second task (gauge task) was to monitor the temperature of the radar displayed on a meter (Figure 2). If the temperature dropped below zero or exceeded 300, participants had to press the control key to reset the meter. Resetting the meter either too soon or too late resulted in a lower score.

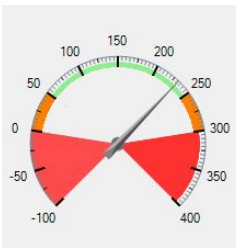

Figure 2. Meter (gauge) that displays the temperature of the radar.

Design

A 3 (model type) * 2 (task complexity) * 2 (performance level) design was used. Task complexity is a within-subjects factor and the order was balanced between the participants. Performance level was a quasi independent variable we used to categorize participants.

Independent variables

Type of cognitive model of attention. The gaze-based model uses gaze behavior to estimate where attention is allocated. Eye-movements of the participants were recorded with an eye-tracker. The task-based model uses information about how the task is expected to be performed. The combined model uses both gaze behavior and the task-based model to determine where attention is allocated (Bosse, van Maanen, & Treur, 2008).

Complexity of the task. A complex and simple scenario were developed by manipulating the ambiguity and the dynamics of the scenario of the tactical picture compilation task. Concerning ambiguity, with small differences in the threat level of contacts it will more difficult to identify the five most threatening contacts. Dynamics was manipulated by varying the number of threat level changes of contacts over time. With many changes in the threat level it will be difficult to identify the five most threatening contacts, because the number of times that the contacts need to be re-evaluated increases.

Performance level. After the experiment, a median split was performed to separate good and poor performers.

Dependent variables

Task performance. The performance on the tactical picture compilation task was determined by the accuracy of the identification of the five most threatening contacts during the task. The performance on the gauge task was determined by the accuracy of resetting the meter.

Accuracy of the models. At random moments, varying from 2–6 minutes, both tasks were frozen. During a freeze, participants had to select the contacts by clicking on those that received their attention in the past 3 seconds. None of the contacts had to be selected when attention was only allocated to the temperature meter. The selected contacts were matched with those predicted by the models. The accuracy of each model was determined by calculating the area under the ROC curve (AUC) per model (Swets, 1988). The AUC indicates how accurate a model is to describe the participant's dynamics of attention allocation.

Procedure

Both tasks were explained thoroughly to the participants before the experiment started. The criteria that had to be used for the tactical picture compilation task were explained using different examples. All participants were tested on whether they were able to correctly apply the criteria: when the score was below 80%, they received extra instructions and another test. Participants performed a practice trial in which they had to perform both tasks. It was stressed that both tasks were equally important and that both tasks had to be performed well to attain a good performance overall.

RESULTS

Manipulation check

The difference in task complexity between a simple and complex scenario was determined by measuring the performance on the tactical picture compilation task. The performance in the simple scenario was significantly higher than the performance in the complex scenario ($t(41) = 4.56$, $p < 0.01$). A median split was performed to the divide the group in good and poor performers. A t-test showed that the difference in performance between these groups was significant ($t(40) = -23.13$, $p < 0.01$).

Main effects

A three-way ANOVA with planned comparisons was performed to test the hypotheses. A significant main effect was found for the model type ($F(2, 39) = 8.97$, $p < .01$), but not for task complexity ($F(1, 40) = 0.29$, $p = .59$) and task performance ($F(1, 40) = 1.35$, $p = .25$).

Effect of model type

The results of the paired right-tailed t-tests for hypothesis 6 are shown in Table 1. As was expected, the AUC of the combined model was significantly higher than that of the task-based and the gaze-based model. Hypothesis 6 is therefore accepted.

Table 1. *T-tests for H6*

Hypothesis	Description	M_1 (SD_1)	M_2 (SD_2)	t	df	p
H6	C > G	.66 (.08)	.58 (.06)	4.98	82	.00*
	C > T	.66 (.08)	.62 (.08)	2.19	82	.02*

Note. C = combined, G = gaze-based, T = task-based model.
**p* < 0.05

Effect of task complexity

Figure 3 displays the differences between the mean AUC for all three models in the simple and complex conditions. The results of the paired right-tailed t-tests for hypotheses 1, 2, and 3 are shown in Table 2.

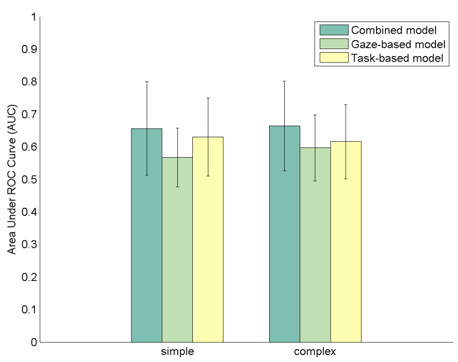

Figure 3. Model performance (AUC): simple and complex condition.

For all three models there were no significant differences in AUC between the simple and complex condition. Hypothesis 1 is therefore rejected. The AUC of the combined model was higher than that of the other models in both the simple and complex condition. However, not all differences were significant. In the simple condition, the difference between the combined model and the task-based model was not significant. Hypothesis 2 is therefore not fully accepted. Furthermore, it was expected that the difference between the combined model and the other models is higher in the complex condition than in the simple condition. This difference was not significant. Hypothesis 3 is not accepted.

Table 2. *T-tests for H1, H2, and H3*

Hypothesis	Description	M_1 (SD_1)	M_2 (SD_2)	t	df	p
	Cs > Cc	.66 (.14)	.66 (.14)	-.28	82	.61
H1	Gs > Gc	.57 (.09)	.57 (.09)	-1.43	82	.92
	Ts > Tc	.63 (.12)	.62 (.11)	.54	82	.30
	Cs > Gs	.66 (.14)	.57 (.09)	3.40	82	.00*
H2	Cs > Ts	.66 (.14)	.63 (.12)	.91	82	.18
	Cc > Gc	.66 (.14)	.60 (.10)	2.57	82	.01*
	Cc > Tc	.66 (.14)	.62 (.11)	1.77	82	.04*
H3	Cc−Gc > Cs−Gs	.07 (.06)	.09 (.09)	-1.33	82	.91
	Cc−Tc > Cs−Ts	.05 (.09)	.03 (.06)	1.41	82	.08

Note. C = combined, G = gaze-based, T= task-based model, s = simple and c = complex condition.
**p* < 0.05

Effect of task performance

Figure 4 displays the differences between the mean AUC for all three models for good and poor performers. The results of the right-tailed t-tests for hypothesis 4 (unpaired) and 5 (paired) are shown in Table 3.

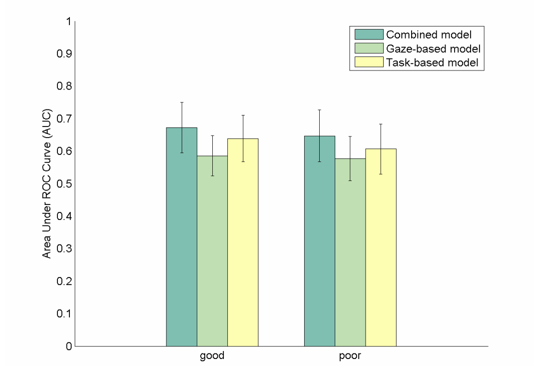

Figure 4. Model performance (AUC): good and poor performer condition.

Table 3. *T-tests for H4 and H5*

Hypothesis	Description	M_1 (SD_1)	M_2 (SD_2)	t	df	p
H4	Cg > Cp	.67 (.08)	.65 (.08)	1.03	40	.15
	Tg > Tp	.64 (.07)	.61 (.08)	1.41	40	.08
	Cg > Gg	.67 (.08)	.59 (.06)	3.96	40	.00*
H5	Cg > Tg	.67 (.08)	.64 (.07)	1.44	40	.08
	Cp > Gp	.65 (.08)	.68 (.07)	3.07	40	.00*
	Cp > Tp	.65 (.08)	.61 (.08)	1.67	40	.05

Note. C = combined, G = gaze-based, T = task-based model, g = good and p= poor performers condition.
* *p* <0.05

Contrary to what was expected, no significant difference was found between good and poor performers for the combined and task-based model. Hypothesis 4 is not accepted. The AUC of the combined model is higher than the AUC of the other two models for both good and poor performers. The difference between the combined and gaze-based model was significant. Hypothesis 5 is therefore partly accepted.

CONCLUSION AND DISCUSSION

In this experiment participants had to execute a naval tactical picture compilation task. During the task participants had to indicate to what set of objects (ships) on the screen their attention was allocated. This was compared with the estimates of three cognitive models of attention: a gaze-based model, a task-based model and a combined model. Model performance was calculated for simple and complex task scenarios and for good and poor performers.

Results show that overall the combined model was a significant better predictor of attention allocation than the gaze-based and the task-based model. For both simple and complex task scenarios and for both good and poor performers the performance of the combined model was better than the performance of the other models. However, these differences were not always significant. In the complex scenario, the combined model was significantly better than both models. In the simple scenario the combined model was only significantly better than the gaze-based model. It seems that in simple tasks the inclusion of gaze-based information to the task-based model does not result in significantly more predictive power. This might be explained by that the complexity of models, e.g. single versus combined models, only has a positive effect when it is applied in complex scenarios. For good and poor performers the combined model was significantly better than the gaze-based model, but not significantly better than the task-based model. Our results indicate that for both good and poor performers, the inclusion of gaze-based information to the task-based model does not result in significantly more predictive power.

We did not find that the models were consistently better in the simple scenario compared to the complex scenario. We also did not find that the combined model and the task-based model were better for good performers than for poor performers. The difference in task complexity and performance level or the number of participants may have been insufficient to fully confirm our hypotheses. Even though the manipulation check showed significant differences in performance of the participants between both conditions, these differences might not have been enough to cause an effect in the performance of the models.

Although it is too early to tell whether the predictions of the combined model are reliable enough for application in support systems, our results suggest that a combined model is the best candidate for this in terms of robustness and performance. To enhance the performance of the models, optimal parameter values need to be determined. Further, more research is needed to determine whether the predictive power of the models can be improved by adding components, such as expertise of the user or the visual saliency of information. Such an investigation can be done using similar validation techniques as described in this paper. Furthermore, since the AUC performance measure is decision criterion independent, it needs to be determined whether liberal or conservative criterion settings are more effective for the prediction of human attention allocation or whether this criterion should be determined dynamically. Finally, in the near future experiments in which participants perform a tactical picture compilation task with and without support based on a combined model are needed as a more objective means for model evaluation.

ACKNOWLEDGMENT

This research was funded by the Dutch Ministry of Defence under program number V524. The authors would like to thank Jan Maarten Schraagen for his helpful comments.

REFERENCES

Bosse, T., Maanen, P.-P. van, Treur, J., Simulation and Formal Analysis of Visual Attention, Journal of Web Intelligence and Agent Systems (WIAS), IOS Press, 2008.

Chalmers, B. A., Webb, R. D. G., & Keeble, R. (2002). Modeling shipboard tactical picture compilation. *Proceedings of the Fifth International Conference on Information Fusion*. Volume 2, Sunnyvale, CA, International Society of Information Fusion, pp. 1292–1299.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *Int. J. Human-Computer studies*, *58,* 697-718.

Grootjen, M., & Neerincx, M. (2005). Operator load management during task execution in process control. *Human Factors Impact on Ship Design*, Genoa.

Just, M., & Carpenter, P. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441-480.

Kahneman, D. (1973). *Attention and effort*. Prentice Hall, Englewoods Cliffs, NJ.

Maanen, P.-P. van, Koning, L. de, Dongen, K. van (2007). Design and Validation of HABTA: Human Attention-Based Task Allocator, In: *Proceedings of the First International Workshop on Human Aspects in Ambient Intelligence*, Darmstadt, Germany, to be published by Springer.

Ouerhani, N., Von Wartburg, R., Hügli, H., & Müri, R. (2004). Emperical Validation of the saliency-based Model of Visual Attention. *Electronic Letters on Computer Vision and Image Analysis*, *3*(1), 13-14.

Parasuraman & Riley (1997). Humans and Automation: Use, Misuse, Disuse, and Abuse. *Human Factors, 39*(2*),* 230-253.

Pavel, M., Wang, G. & Li, K. (2003). Augmented cognition: Allocation of attention. *Proceedings of the 36th Hawaii International Conference on System Sciences* (HICSS'03).

Salvucci, D.D. (2000). A model of eye movements and visual attention. *Third International Conference on Cognitive modeling*, pp. 252-259.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. Science, *240*, 1285–1293.

Treisman & Gelade (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Verkuijlen, R. P. M., Muller, T. J. (2007). *Action speed tactical trainer review*. Technical report, TNO Human Factors.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention,* 63–102. San Diego, CA: Academic Press.

Wickens, C. D., & McCarley, J. S. (2007). *Applied attention theory*. CRC Press.

Wood, R.E. (1986). Task complexity: definition of construct. *Organizational Behavior and Human Decision Processes*, *37*, 60-82.

Zijlstra, F. R. H., & Doorn, L.V. (1985). *The Construction of a Scale to Measure Perceived Effort.* Department of Philosophy and Social Sciences, Delft University of Technology, Delft, The Netherlands.