

Automated Visual Attention Manipulation

Tibor Bosse¹, Rianne van Lambalgen¹, Peter-Paul van Maanen^{1,2}, and Jan Treur¹

¹ Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
{tbosse, rm.van.lambalgen, treur}@cs.vu.nl

² TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
peter-paul.vanmaanen@tno.nl

Abstract. In this paper a system for visual attention manipulation is introduced and formally described. This system is part of the design of a software agent that supports naval crew in her task to compile a tactical picture of the situation in the field. A case study is described in which the system is used to manipulate a human subject's attention. To this end the system includes a Theory of Mind about human attention and uses this to estimate the subject's current attention, and to determine how features of displayed objects have to be adjusted to make the attention shift in a desired direction. Manipulation of attention is done by adjusting illumination according to the calculated difference between a model describing the subject's attention and a model prescribing it.

1 Introduction

In the domain of naval warfare, it is crucial for the crew of the vessels involved to be aware of the situation in the field. Examples of important questions that should be addressed continuously are “in which direction are we heading?”, “are we currently under attack?”, “are there any friendly vessels around?”, and so on. To assess such issues, one of the crew members is usually assigned the Tactical Picture Compilation Task (TPCT): the task to identify and classify all entities in the environment (e.g., [11]). This is done by monitoring a radar screen for radar contacts, and reasoning with the available information in order to determine the type and intent of the contacts on the screen. However, due to the complex and dynamic nature of the environment, this person has to deal with a large number of tasks in parallel. Often the radar contacts are simply too numerous and dynamic to be adequately monitored by a single human, which compromises the performance of the task.

For these reasons, it may be useful to offer the human some support from an intelligent ambient system, consisting of software agents that assist him in the execution of the Tactical Picture Compilation Task. For example, in case the human is directing its attention on the left part of a radar screen, but ignores an important contact that just entered the radar screen from the right, such an agent may alert him about the arrival of that new contact. To be able to provide this kind of intelligent support, the system somehow needs to maintain a model of the cognitive state of the human: in this case the human's focus of attention. It should have the capability to attribute mental, and in particular attentional (e.g., [12], [13], [14]) states to the

human, and to reason about these. In psychology and philosophy this characteristic is often referred to as Theory of Mind (or ToM, see, e.g., [1]). According to [7], agents, both human and software, can exploit a Theory of Mind for two purposes: to *anticipate* the behaviour of other agents (e.g., preparing for the consequences of certain actions that the other will probably perform), and to *manipulate* it (e.g., trying to influence the actions that the other will perform). In case of an intelligent system to support naval crew members, both purposes are relevant, but require a different type of support. This study is related to the latter type, the type that tries to manipulate the focus of attention.

A number of approaches in the literature address the development of software agents with a Theory of Mind; e.g., [16], [7]. Usually, such agents maintain, in one way or the other, a model of the epistemic (e.g., beliefs) and/or motivational states (e.g., desires, intentions) of other agents. However, for the situation sketched above, such agents ideally also have insight in another agent's attentional states. After all, if a supportive agent is to find out whether the human is ignoring some contact, it needs to have some knowledge about which contacts the person is paying attention to. This idea is in line with the theories of cognitive scientists like Gärdenfors [9], [10], who claims that humans have a Theory of Mind that is not only about beliefs, desires, and intentions, but also about other mental states like attentional, emotional, and awareness states.

The current paper is the result of a project that aims to develop intelligent agents to support naval crew members in the Tactical Picture Compilation Task, based on the ideas described above. To this end, four models have been developed. First, a dynamical model of human attention is needed, which estimates where the person's attention is, based on information about features of objects on the screen and the person's gaze. Second, a reasoning model is needed to reason through the first model in order to generate beliefs on attentional states at any point in time. Third, a model is needed that compares the output of the second model with some normative attention distribution, and determines whether there is a discrepancy. Finally, a model is needed that uses the output of the third model to determine how to alert the human that he is ignoring something important. An initial version of the first two models has already been developed and were adopted from this earlier work ([5], [6], respectively [2]). The current paper has its focus on the development of the other two models.

Section 2 presents a brief introduction of the existing literature on attentional processes, which helps to understand the choices made within this paper. Next, Section 3 formally describes the different models within the supportive software agent, and presents some simulations that were performed to test the behaviour of the model at a conceptual level. In Section 4, the whole approach is applied in a real-world case study, using human gaze data and a tactical picture compilation task environment. Finally, Section 5 is a discussion.

2 Manipulation of Attention

Typically, a person's attention is influenced both by top-down and by bottom-up processes. The former means that observers orient their attention in a goal-directed manner, as a consequence of their expectations or intentions [19]. For example, when

searching for a friend in the crowd, attention is guided top-down [20]. In contrast, the latter means that attention is elicited by a (highly salient) trigger from the environment. For example, one green circle among several blue circles will “pop-out” and attention will be directed to this object [22].

In this project the focus is primarily on adjusting the features of a specific location, such that only bottom-up attention is manipulated. Features that are mainly known to influence attention are intensity (luminance), colour and orientation. Previous research shows that attention can be elicited both by the contrast with stimuli at other locations [12], [15], [17] and the abrupt change of a feature, like luminance [21], [23] or form [23].

Several cognitive models on attention have been proposed and show that it is possible to predict attention allocation based on a saliency map, calculated from features of a stimulus, like luminance, colour and orientation [13], [18]. These models are not dynamic in the sense that they are based on existing information from the environment. However, if indeed the change of a specific feature (like luminance) can cause an attention shift in the human performing a task considered, a model can be used to realise this change. This way, humans who have to direct their attention to a large number of locations in parallel can be supported to adequately perform their task.

3 Formalisation of a Theory of Mind for Attention

In this section it is shown how the Theory of Mind for attention within the software agent was designed. First, in Section 3.1 the general setting is described, distinguishing four models. In subsequent subsections 3.2, 3.3, 3.4, and 3.5 these four models are described in more detail.

3.1 Overall Setting

A Theory of Mind enables an agent to analyze another agent’s mind, and to act according to the outcomes of such an analysis and its own goals. For the general case such processes require some specific facilities.

(1) A *representation of a dynamical model* is needed describing the relationships between different mental states of the other agent. Such a model may be based on qualitative causal relations, but it may also concern a numerical dynamical system model that includes quantitative relationships between the other agent’s mental states. In general such a model does not cover all possible mental states of the other agent, but focuses on certain aspects, for example on beliefs and desires, on emotional states, on the other agent’s awareness states, or on attentional states as in this paper.

(2) Furthermore, *reasoning methods to generate beliefs on the other agent’s mental state* are needed to draw conclusions based on the dynamical model in (1) and partial information about the other agent’s mental states. This may concern deductive-style reasoning methods performing forms of simulation based on known inputs to predict certain output, but also abductive-style methods reasoning from output of the model to (possible) inputs that would explain such output.

(3) Moreover, when in one way or the other an estimation of the other agent’s mental state has been found out, it has to be *assessed whether there are discrepancies*

between this state and the agent's own goals. Here also the agent's self-interest comes in the play. It is analyzed in how far the other agent's mental state is in line with the agent's own goals, or whether a serious threat exists that the other agent will act against the agent's own goals.

(4) Finally a *decision reasoning model* is needed to decide how to act on the basis of all of this information. Two types of approaches are possible. A first approach is to take the other agent's state for granted and prepare for the consequences to compensate for them as far as these are in conflict with the agent's own goals, and to cash them as far as they can contribute to the agent's own goals (*anticipation*). For the navy case, an example of anticipation is when it is found out that the other agent has no attention for a dangerous object, and it is decided that another colleague or computer system will handle it (dynamic task reallocation). A second approach is not to take the other agent's mental state for granted but to decide to try to get it adjusted by affecting the other agent, in order to obtain a mental state of the other agent that is more in line with the agent's own goals (*manipulation*). This is the case addressed in this paper.

In this paper the general pattern sketched above is applied to the way in which a (software) agent can attempt to adjust the other (human) agent's attention, whenever required. To this end the software agent uses four types of facilities:

- *A dynamical model for attention*
Representation of a dynamical attention model: a model that provides as output an estimation of the current attention distribution, based on input about features of objects on the screen and the other agent's gaze.
- *A reasoning model to generate beliefs about attentional states*
These methods are used to estimate the attention given inputs about features of the objects and the other agent's gaze.
- *A discrepancy assessment model*
This concerns a model to determine whether it is desirable that the attention distribution is changed, and to which extent: the discrepancy between actual and desirable attentional states
- *A decision reasoning model*
This is a model to determine how, given a desire to adjust the attention distribution in certain respects, the inputs for the attention model have to be changed, to obtain an attention distribution as output, which is adjusted as desired.

The dynamical attention model is taken over from [5], [6] and is only briefly summarised below. The second model is kept relatively simple: beliefs on the attentional state are generated just based on (internal) simulation of the attention model. The third and fourth model form the most crucial part of this paper.

3.2 The Dynamical Attention Model Used

The attention distribution at *time* t is an assignment of attention values $AV(s, t)$ to a set of attention *spaces* s . Attention spaces are squares within a grid. The attention

distribution is assumed to have a certain persistency. At each point in time the new attention is related to the previous attention, by

$$AV(s, t) = \lambda \cdot AV(s, t - 1) + (1 - \lambda) \cdot AV_{\text{norm}}(s, t)$$

where λ is the decay parameter that results in the decay of the attention value of space s at time point $t - 1$. Note that higher values for λ result in a higher persistency and lower decay and vice versa. Here $AV_{\text{norm}}(s, t)$ is determined by a normalisation process keeping the total amount of attention fixed. This is described by:

$$AV_{\text{norm}}(s, t) = \frac{AV_{\text{new}}(s, t)}{\sum_{s'} AV_{\text{new}}(s', t)} \cdot A(t)$$

$$AV_{\text{new}}(s, t) = \frac{AV_{\text{pot}}(s, t)}{1 + \alpha \cdot r(s, t)^2}$$

Here $AV_{\text{new}}(s, t)$ is defined as follows. An important aspect of the visual attentional state is human gaze behaviour. Therefore the relative distance of each space s to the gaze point (the centre) is an important factor in determining the attention value of s . Mathematically this is modelled by the formula above, where $AV_{\text{pot}}(s, t)$ is the potential attention value of s at time point t . The term $r(s, t)$ is taken as the Euclidian distance between the current gaze point and s at time point t (multiplied by an importance factor α which determines the relative impact of the distance to the gaze point on the attentional state, which can be different per individual and situation):

$$r(s, t) = d_{\text{eucl}}(\text{gaze}(t), s)$$

$$AV_{\text{pot}}(s, t) = \sum_{\text{maps } M} M(s, t) \cdot w_M(s, t)$$

Here the potential attention value $AV_{\text{pot}}(s, t)$ is calculated as follows, based on the properties of the space (i.e., of the types of objects present) at that time (for instance features such as colour, intensity, and orientation contrast, amount of movement). For each of such a feature a specific *saliency map* describes its potency of drawing attention; e.g., [8], [13], [14]. Because not all features are equally highlighting, an additional weight for every map is used. Formally the above can be described as shown above, where for any feature there is a saliency map M , for which $M(s, t)$ is the unweighted potential attention value of s at time point t , and $w_M(s, t)$ is the weight for saliency map M , where $1 \leq M(s, t)$ and $0 \leq w_M(s, t) \leq 1$. The exact values for the weights depend on the specific application.

3.3 Reasoning Model to Generate Beliefs About Attentional States

The reasoning method to generate beliefs about attentional states is kept simple. The model described in Section 3.2 as a dynamical system model (based on a difference equation) is just used by the software agent as an internal simulation model to generate new attentional states out of the previous ones and information about the current features of the objects. This is done by a forward reasoning method (forward in time) as described in [2]. This reasoning method can be used to make predictions on future states, or on making an estimation of the current state based on information

acquired in the past. This reasoning method occurs in the literature in many variants, in different contexts and under different names, varying from, for example, computational (numerical) simulation based on difference or differential equations, qualitative simulation, causal reasoning, execution of executable temporal logic formulae, and forward chaining in rule-based reasoning, to generation of traces by transition systems and finite automata. The basic specification of this reasoning model can be expressed as follows, where $\text{belief}(\text{leads_to_after}(I, J, D))$ (the belief that when state I holds at time T , then J will hold after time duration D) is used as the agent's internal representation format for dynamical system models, and $\text{belief}(\text{at}(I, T))$ as a representation of its information on the world (including human processes) at different points in time; moreover, \rightarrow means that when the antecedent holds, the consequent will follow. Here, I and J are predicates that may represent world states like 'AV(s,t) = 0.5' or 'luminance value 0.8 is assigned to s'.

Belief Generation based on Positive Forward Simulation

If it is believed that I holds at T and that I leads to J after duration D , then it is believed that J holds after D .

$$\text{belief}(\text{at}(I, T)) \wedge \text{belief}(\text{leads_to_after}(I, J, D)) \rightarrow \text{belief}(\text{at}(J, T+D))$$

If it is believed that I_1 holds at T and that I_2 holds at T , then it is believed that I_1 and I_2 holds at T .

$$\text{belief}(\text{at}(X_1, T)) \wedge \text{belief}(\text{at}(X_2, T)) \rightarrow \text{belief}(\text{at}(\text{and}(X_1, X_2), T))$$

This is done by calculations following the formulae described above.

3.4 A Model to Determine Discrepancy between Actual and Desirable Attention

The discrepancy between actual and desirable attention can be determined as soon as a model is available for what the desirable attention distribution is (sometimes called a prescriptive model). For the case addressed in this paper this means that in a computational manner it is assessed which objects deserve attention, an assessment on the basis of features such as distance, speed and direction of an object¹. In fact, this is close to the first part of the task the human is performing: identification of the relevant objects to be handled.

3.5 A Decision Model for Attention Adjustment

The model for adjustment of the attention distribution has as input the discrepancy determined by the model described in Section 3.4, and also makes use of the explicitly represented dynamical model as described in Section 3.2. The general idea is that the relations between variables within this model are followed in a backward manner, thereby propagating the desired adjustment from the attentional state variable to the features of the object at the screen. The general pattern behind this operation on a dynamical model representation is illustrated in Figure 1. Here v_j is the (desired)

¹ It is assumed that the agent has tactical domain knowledge that enables it to make such assessments.

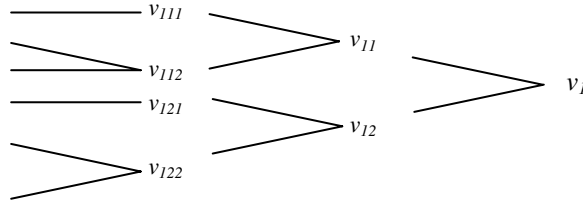


Fig. 1. Dependencies between variables in a dynamical system model

output of a model, and by branches the variables on which this depends are depicted, until the leaves where actual adjustments can be made².

This is a form of *desire refinement*: starting from the root variable, by a step-by-step process a desire on adjusting a parent variable is refined to desires on adjustments of the children variables, until the leaf variables are reached. The starting point is the desire on the root variable, which is the desired adjustment of the attentional state; this is determined by.

$$\text{belief}(av(s)<h) \wedge \text{desire}(a(v)>h) \wedge \text{belief}(\text{has_value}(av(s), v)) \rightarrow \text{desire}(\text{adjust_by}(av(s), (h-v)/v))$$

Note that here the adjustment is taken relative (expressed by division of the difference h-v by v). Suppose as a point of departure (given the discrepancy assessment) an adjustment Δv_1 is desired, and that v_1 depends on two variables v_{11} and v_{12} that are adjustable (the non-adjustable variables can be left out of consideration). Then by elementary calculus as a linear approximation the following relations between required adjustments can be obtained:

$$\Delta v_1 = \frac{\partial v_1}{\partial v_{11}} \Delta v_{11} + \frac{\partial v_1}{\partial v_{12}} \Delta v_{12}$$

This formula is used to determine the desired adjustments Δv_{11} and Δv_{12} , where by weight factors μ_{11} and μ_{12} the proportion can be indicated in which the variables should contribute to the adjustment: $\Delta v_{11}/\Delta v_{12} = \mu_{11}/\mu_{12}$.

$$\Delta v_1 = \frac{\partial v_1}{\partial v_{11}} \Delta v_{12} \mu_{11}/\mu_{12} + \frac{\partial v_1}{\partial v_{12}} \Delta v_{12} = \left(\frac{\partial v_1}{\partial v_{11}} \mu_{11}/\mu_{12} + \frac{\partial v_1}{\partial v_{12}} \right) \Delta v_{12}$$

So the adjustments can be made as follows:

$$\Delta v_{12} = \frac{\Delta v_1}{\frac{\partial v_1}{\partial v_{11}} \mu_{11}/\mu_{12} + \frac{\partial v_1}{\partial v_{12}}}$$

$$\Delta v_{11} = \mu_{11}/\mu_{12} \frac{\Delta v_1}{\frac{\partial v_1}{\partial v_{11}} \mu_{11}/\mu_{12} + \frac{\partial v_1}{\partial v_{12}}} = \frac{\Delta v_1}{\frac{\partial v_1}{\partial v_{11}} + \frac{\partial v_1}{\partial v_{12}} \mu_{12}/\mu_{11}}$$

Special cases are $\mu_{11} = \mu_{12} = 1$ (*absolute equal contribution*) or $\mu_{11} = v_{11}$ and $\mu_{12} = v_{12}$ (*relative equal contribution*: in proportion with their absolute values). As an

² For the moment, deterministic relationships between variables are assumed. However, in a later stage, the agent might learn such relationships.

example, consider a variable that is just the weighted sum of two other variables (as is the case, for example, for the aggregation of the effects of the features of the objects on the attentional state):

$$v_I = w_{I1}v_{I1} + w_{I2}v_{I2}$$

For this case

$$\frac{\partial v_I}{\partial v_{I1}} = w_{I1} \qquad \frac{\partial v_I}{\partial v_{I2}} = w_{I2}$$

and

$$\Delta v_{I1} = \frac{\Delta v_I}{w_{I1} + w_{I2} \mu_{I2}/\mu_{I1}} \qquad \Delta v_{I2} = \frac{\Delta v_I}{w_{I1} \mu_{I1}/\mu_{I2} + w_{I2}}$$

For example when $\mu_{I1} = \mu_{I2} = 1$ this results in

$$\Delta v_{I1} = \frac{\Delta v_I}{w_{I1} + w_{I2}} \qquad \Delta v_{I2} = \frac{\Delta v_I}{w_{I1} + w_{I2}}$$

Assuming $w_{I1} + w_{I2} = 1$ in addition, this results in $\Delta v_{I1} = \Delta v_{I2} = \Delta v_I$

Another setting, which actually has been used in the model is to take $\mu_{I1} = v_{I1}$ and $\mu_{I2} = v_{I2}$. In this case the adjustments are assigned proportionally; for example, when v_I has to be adjusted by 5%, also the other two variables on which it depends need to contribute an adjustment of 5%. Thus the relative adjustment remains the same through propagations:

$$\frac{\Delta v_{I1}}{v_{I1}} = \frac{\Delta v_I}{w_{I1} + w_{I2} v_{I2}/v_{I1}} / v_{I1} = \frac{\Delta v_I}{w_{I1} v_{I1} + w_{I2} v_{I2}} = \frac{\Delta v_I}{v_I}$$

This shows the general approach on how desired adjustments can be propagated in a backward manner through a dynamical model. Thus a desired adjustment of the attentional state as output at some point in time can be related to adjustments in the features of the displayed objects as inputs at previous points in time. For the case study undertaken this approach has been applied, although at some points in a simplified form. One of the simplifications made is that due to the linearity of most dependencies in the model, adjustments have been used that just propagate without any modification. An example of a rule specified to achieve this propagation process is:

$$\text{desire}(\text{adjust_by}(u1, a)) \wedge \text{belief}(\text{depends_on}(u1, u2)) \rightarrow \text{desire}(\text{adjust_by}(u2, a))$$

Here the adjustments are taken relative, so, this rule is based on $\Delta u_2 / u_2 = \Delta u_1 / u_1$ as derived above for the linear case. When at the end the leaves are reached, which is represented by the belief that they are directly adjustable, then from the desire an intention to adjust them is derived.

$$\text{desire}(\text{adjust_by}(u, a)) \wedge \text{belief}(\text{directly_adjustable}(u)) \rightarrow \text{intention}(\text{adjust_by}(u, a))$$

If an intention to adjust a variable u by a exists with current value b , the new value $b + \alpha \cdot a \cdot b$ to be assigned to u is determined; here α is a parameter that allows the modeler to tune the speed of adjustment:

$$\text{intention}(\text{adjust_by}(u, a)) \wedge \text{belief}(\text{has_value_for}(u, b)) \rightarrow \text{performed}(\text{assign_new_value_for}(u, b + \alpha * a * b))$$

This rule is applied for variables that describe features f of objects at locations s , i.e., instances for u of the form $\text{feature}(s, f)$. Note that each time the adjustment is propagated as a value relative to the overall value.

3.6 Simulation Results

To test whether the approach described above yields the expected behaviour, it has been used to perform a number of simulation experiments in the LEADSTO simulation environment [4]. This environment takes a specification of causal relationships (in the format as shown in the previous sections) as input, and uses this to generate simulation traces. The simulations shown here address a slightly simplified case, where the radar screen has been split up in 4 locations. For the time being, it is assumed that each location contains one contact, and that these contacts stay within their locations. The features of the contacts that are manipulated are luminance, size, and level of flashing. Initially, each contact starts with the same features, but during the simulation these features are manipulated, based on the prescribed (or desired) attention. This desired attention is generated randomly, where every 50 time units a next location is selected where the attention should be. Furthermore, the behaviour of the human gaze is generated as follows: after each adaptation of the features, the gaze moves to one of the four locations, with a probability that is proportional to the saliency of the contact at that location.

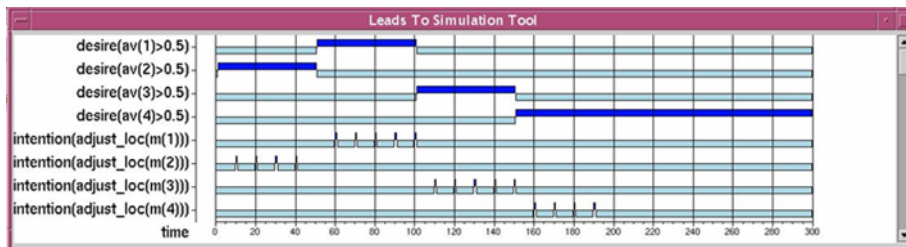


Fig. 2. Model-based reasoning process. First it is intended (several times) to adjust a feature value at location 2, then at location 1, then at location 3, and finally at location 4.

The results of an example simulation run are depicted in Figures 2 to 5. In these figures, time is on the horizontal axis, and the different state of the process is shown in the vertical axis. A dark line indicates that a state is true at a certain time point. Note that some information has been omitted due to space limitations. Figure 2 shows the model-based reasoning process of the agent, in terms of desires and intentions. Figures 3, 4, and 5 show, respectively, the estimated attention, the human's gaze, and the value of the feature "luminance" at different locations over time. As shown in Figure 2, initially it is desired that at least 50% of the human's attention is at location 2 ($\text{desire}(\text{av}(2) > 0.5)$). Since this is not the case (see Figure 3), the luminance of the

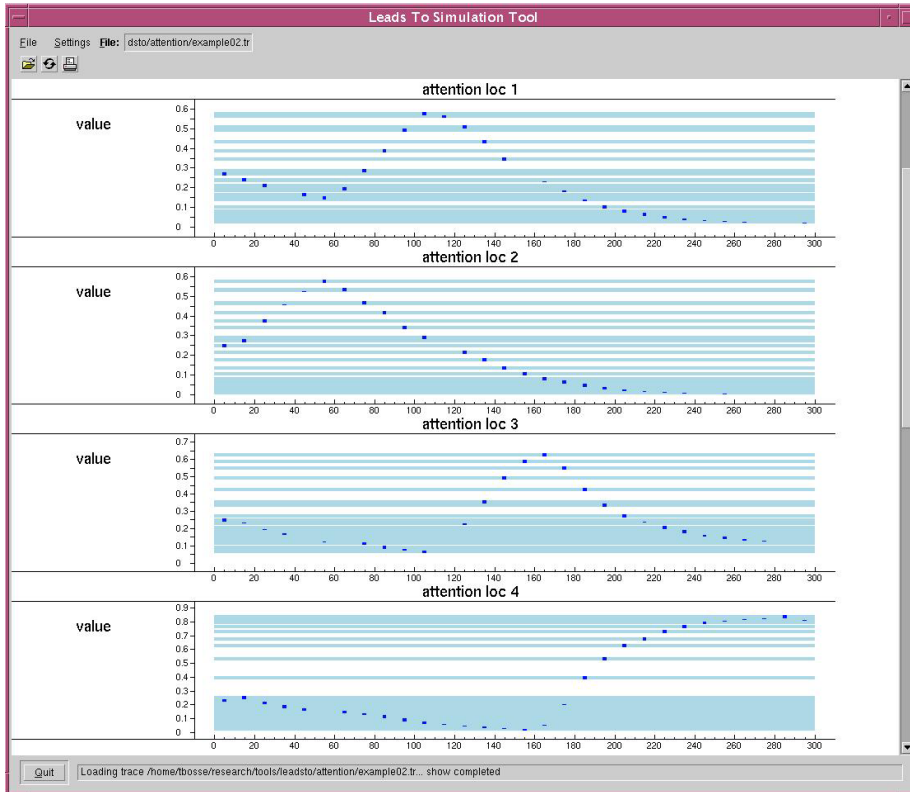


Fig. 3. Estimated attention at different locations. Initially the highest attention value is estimated to be at location 2 (with a peak around time point 55), then at location 1, then at location 3, and finally at location 4.

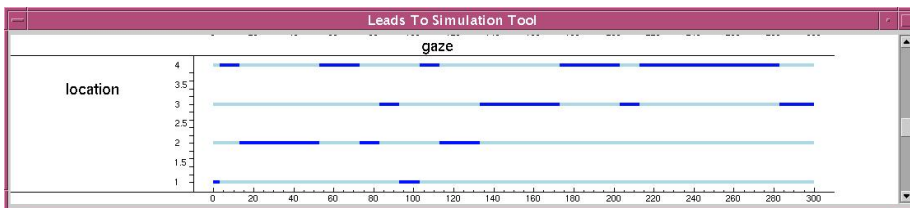


Fig. 4. Dynamics of gaze. The vertical axis denotes the location of the gaze, which switches between location 1, 2, 3, and 4.

contact at location 2 is increased (see Figure 5). As a result, the human’s gaze shifts towards this location (see Figure 4), which increases his attention for location 2. In the rest of the simulation, this pattern is repeated for different locations.

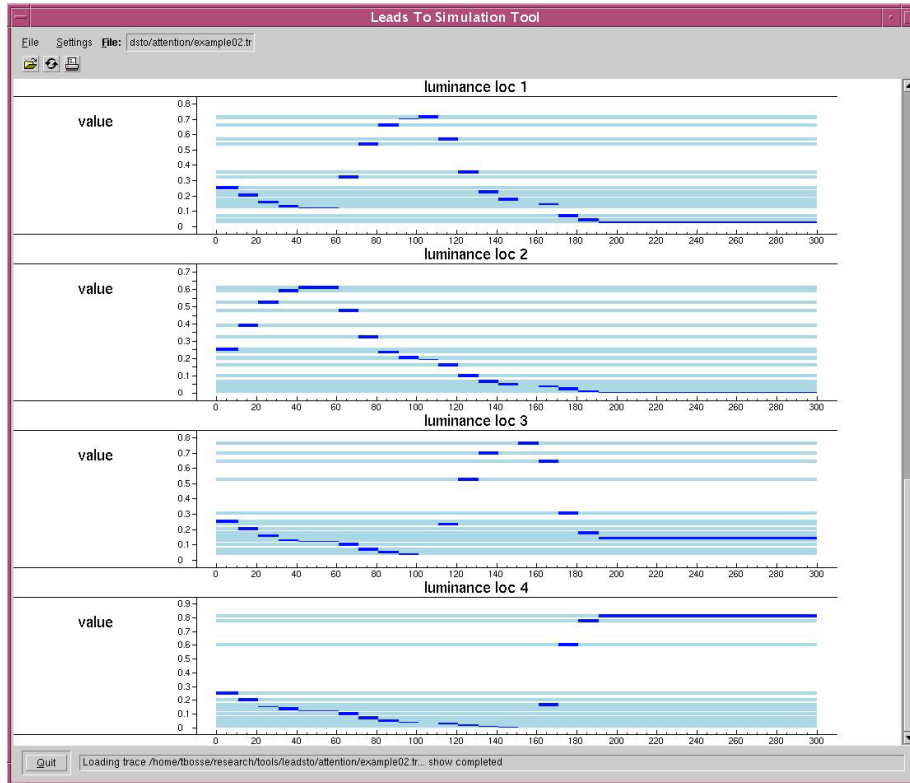


Fig. 5. Values of feature ‘luminance’ at different locations. First the luminance at location 2 is increased, then at location 1, 3, and 4 (note that values are normalised).

4 Case Study

To test the approach in a real world situation, and obtain an initial validation, a case study with human subjects while executing the Tactical Picture Compilation Task was undertaken. In Section 4.1 the environment is explained, Section 4.2 discusses some implementation details of the system tailored to the environment, and in Section 4.3 the first results are discussed.

4.1 Environment

The task used for this case study is an altered version of the identification task described in [11] that has to be executed in order to build up a tactical picture of the situation, i.e. the Tactical Picture Compilation Task (TPCT). The implementation of the software is done in Gamemaker [25]. In Figure 6 a snapshot of the interface of the task environment is shown. The goal is to identify the five most threatening contacts (ships). In order to do this, participants have to monitor a radar display where contacts in the surrounding areas are displayed. To determine if a contact is a possible threat,

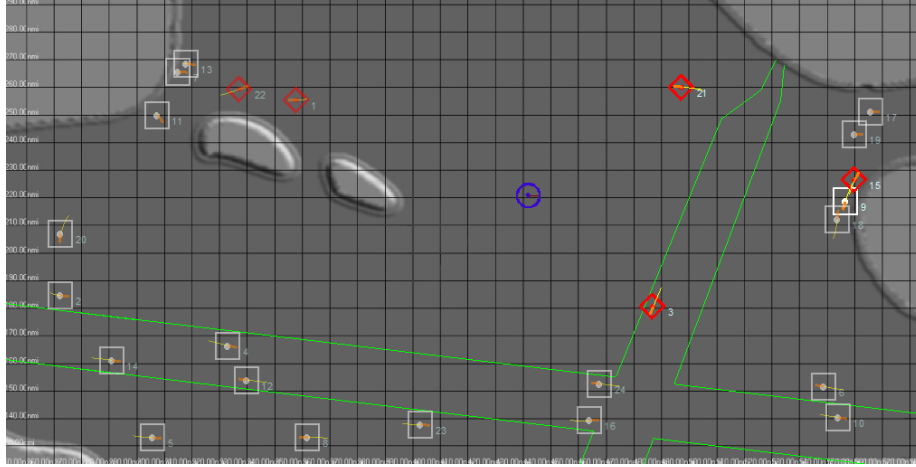


Fig. 6. A snapshot of the interface of the used task environment

different criteria have to be used. These criteria are the identification criteria (idcrits) that are also used in naval warfare, but are simplified in order to let naive participants learn them more easily. These simplified criteria are the speed (depicted by the length of the tail of a contact), direction (pointer in front of a contact), distance of a contact to the own ship (circular object), and whether the contact is in a sea lane or not (in or out the large open cross). Contacts can be identified as either a threat (diamond) or no threat (square).

4.2 Implementation

The system is further developed and evaluated using Matlab. The output of the environment described in Section 4.1 was used and consisted of a representation of all properties of the contacts visible on the screen, i.e. speed, direction, if it is in a sea lane or not, distance to the own ship, location on the screen and contact number. In addition, data from a Tobii x50 eye-tracker [24] were retrieved from a participant executing the TPC task. All data were retrieved several times per second and were used as input for the system. Once the system was tailored to the TPC case study, the eventual implementation of it was done in C#. The output of the implementation of the system causes the saliency of the different objects on the screen to either increase or decrease, which may result in a shift of the participant's visual attention. As a result, the participant's attention is continuously manipulated in such a way that it pays attention to the objects that are considered relevant by the system. The results of this implementation are described below.

4.3 Results

The first results of the implemented system are best described by a number of example snapshots of the outcomes of the system in three different situations over time (see Figure 7).

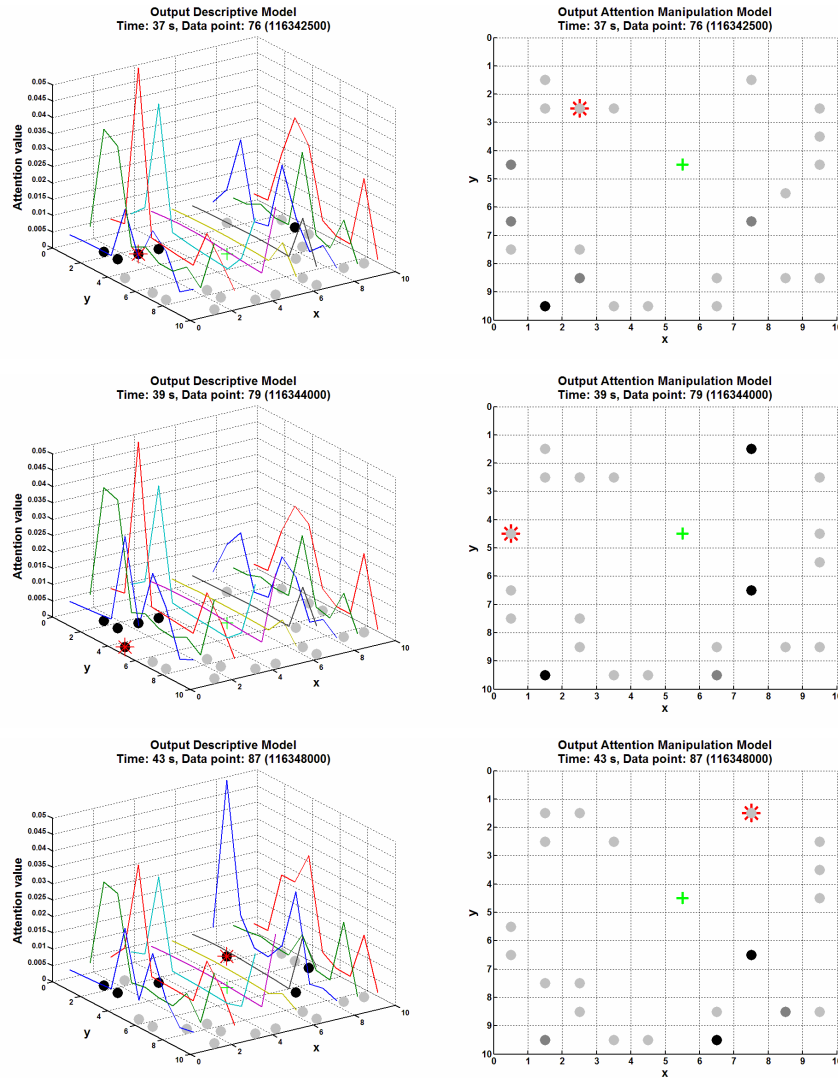


Fig. 7. Estimation of the participant's attention division (left figures) and reaction of the system (right figures) in three different situations

On the left side of Figure 7 the darker dots correspond to the system's estimation of those contacts to which the participant is paying attention. On the right side of the figure, the darker dots correspond to those contacts where attention manipulation is initiated by the system (in this case, by increasing its saliency). On both sides of the figure a cross corresponds to the own ship, a star corresponds to the eye point of gaze, and the x- and y-axes represent the coordinates on the interface of the TPCT. In the pictures to the left, the z-axis represents the estimated amount of attention. The darker dots on the left side are a result of the exceedance of this estimation of a certain

threshold (in this case .03). Thus, a peak indicates that it is estimated that the participant has attention for that location. Furthermore, from top to bottom, the following three situations are displayed in Figure 7:

1. After 37 seconds since the beginning of the experiment, the participant is not paying attention to region A at coordinates (7.5,1.5), while no attention manipulation for region A is initiated by the system.
2. After 39 seconds, the participant is not paying attention to region A, while the attention should be allocated to region A, and therefore attention manipulation for region A is initiated by the system.
3. After 43 seconds, the participant is paying attention to region A, while no attention manipulation for region A is done by the system, because this is not needed anymore.

The output of the attention manipulation system and the resulting reaction in terms of the allocation of the participant's attention in the above three situations, show what one would expect of an accurate system of attention manipulation. As shown in the two pictures at the bottom of Figure 7, in this case the agent indeed succeeds in attracting the attention of the participant: both the gaze (the star in the bottom right picture) and the estimated attention (the peak in the bottom left picture) shift towards the location that has been manipulated.

5 Discussion

An important task in the domain of naval warfare is the Tactical Picture Compilation Task: the task to identify and classify all entities in the environment and determine the consequences in terms of tactical possibilities and constraints. However, due to the complex and dynamic nature of the environment, it is very difficult for a single human to perform this task adequately. Therefore, the current paper proposes to offer the person some support from an intelligent software agent that assists him in the Tactical Picture Compilation Task (TPCT). To this end, a number of models have been developed: 1) a dynamical system model for attention, 2) a reasoning model to generate beliefs about attentional states using the attention model for forward simulation, 3) a discrepancy assessment model, and 4) a decision reasoning model, again using the attention model, this time for backward desire propagation. This paper presented an initial version of such a supporting agent, focusing especially on the last two models, where the first two were adopted from earlier work in [5], [6], and [2]. This software agent is an example of a model-based intelligent ambient agent, as described in [3]. Within this type of agent an explicitly represented (dynamical system) model of human functioning plays an important role, for the case considered here the model of the human's attention. Such a model forms a basis for the application of dedicated model-based reasoning methods, as was also illustrated here: forward simulation reasoning and backward desire propagation reasoning.

After testing the system at a conceptual level by simulation, it has been implemented in a case study where participants have to perform a simplified version of the TPCT. Although no elaborated experimental validation has been performed as

yet, initial results indicate that the agent is indeed able to adapt the features of objects in such a way that they attract the human's attention if necessary.

Concerning future work, an important challenge would be to perform a more elaborated validation of the supportive system. This can be done in several steps. First, to obtain more data, the experiment introduced in this paper will be performed with a larger number of participants. The resulting data can then be used to check (possibly using automated analysis tools) whether the supporting agent is successful in various situations. As part of this validation, also different strategies and parameter settings will be tested. For example, does adapting the shape of an object provide better results than adapting its luminance, or adapting multiple features? Similarly, in addition to manipulation of bottom-up attention, is it useful to manipulate top-down attention as well? Furthermore, in a later stage of the project, it is planned to evaluate whether the software agent indeed improves the task performance of the user.

Acknowledgments

This research was partly funded by the Royal Netherlands Navy (program number V524).

References

- [1] Baron-Cohen, S.: *Mindblindness: an essay on autism and theory of mind*. MIT Press, Cambridge (1995)
- [2] Bosse, T., Both, F., Gerritsen, C., Hoogendoorn, M., Treur, J.: Model-Based Reasoning Methods within an Ambient Intelligent Agent Model. In: Mühlhäuser, M., Ferscha, A., Aitenbichler, E. (eds.) *Proceedings of the First International Workshop on Human Aspects in Ambient Intelligence, Constructing Ambient Intelligence: AmI 2007 Workshops Proceedings*. LNCS. Springer, Heidelberg (2008)
- [3] Bosse, T., Hoogendoorn, M., Klein, M., Treur, J.: An Agent-Based Generic Model for Human-Like Ambience. In: Mühlhäuser, M., Ferscha, A., Aitenbichler, E. (eds.) *Proceedings of the First International Workshop on Model Driven Software Engineering for Ambient Intelligence Applications, Constructing Ambient Intelligence: AmI 2007 Workshops Proceedings*. LNCS. Springer, Heidelberg (to appear, 2008)
- [4] Bosse, T., Jonker, C.M., van der Meij, L., Treur, J.: A Language and Environment for Analysis of Dynamics by Simulation. *International Journal of Artificial Intelligence Tools* 16(3), 435–464 (2007)
- [5] Bosse, T., van Maanen, P.-P., Treur, J.: A Cognitive Model for Visual Attention and its Application. In: Nishida, T. (ed.) *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2006)*, pp. 255–262. IEEE Computer Society Press, Hong Kong (2006)
- [6] Bosse, T., van Maanen, P.-P., Treur, J.: Simulation and formal analysis of visual attention in cognitive systems. In: Paletta, L., Rome, E. (eds.) *WAPCV 2007*. LNCS (LNAI), vol. 4840, pp. 463–480. Springer, Heidelberg (2007)
- [7] Bosse, T., Memon, Z.A., Treur, J.: A Two-Level BDI-Agent Model for Theory of Mind and its Use in Social Manipulation. In: *Proceedings of the AISB 2007 Workshop on Mindful Environments*, pp. 335–342 (2007)
- [8] Chen, L.Q., Xie, X., Fan, X., Ma, W.Y., Zhang, H.J., Zhou, H.Q.: A visual attention model for adapting images on small displays. *ACM Multimedia Systems Journal* (2003)

- [9] Gärdenfors, P.: Slicing the Theory of Mind. In: Danish yearbook for philosophy, vol. 36, pp. 7–34. Museum Tusulanum Press (2001)
- [10] Gärdenfors, P.: How Homo Became Sapiens: On the Evolution of Thinking. Oxford University Press, Oxford (2003)
- [11] Heuvelink, A., Both, F.: Boa: A cognitive tactical picture compilation agent. In: Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2007). IEEE Computer Society Press, Los Alamitos (2007) (forthcoming)
- [12] Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research* 40, 1489–1506 (2000)
- [13] Itti, L., Koch, C.: Computational Modeling of Visual Attention. *Nature Reviews Neuroscience* 2(3), 194–203 (2001)
- [14] Itti, L., Koch, U., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 1254–1259 (1998)
- [15] Levitt, J.B., Lund, J.S.: Contrast dependence of contextual effects in primate visual cortex. *Nature* 387, 73–76 (1997)
- [16] Marsella, S.C., Pynadath, D.V., Read, S.J.: PsychSim: Agent-based modeling of social interaction and influence. In: Lovett, M., Schunn, C.D., Lebiere, C., Munro, P. (eds.) Proc. of the Int. Conference on Cognitive Modeling, ICCM 2004, pp. 243–248 (2004)
- [17] Nothdurft, H.: Saliency from feature contrast: additivity across dimensions. *Vision Research* 40, 1183–1201 (2000)
- [18] Parkurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. *Vision Research* 42, 107–123 (2002)
- [19] Posner, M.E.: Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25 (1980)
- [20] Theeuwes, J.: Endogenous and exogenous control of visual selection. *Perception* 23, 429–440 (1994)
- [21] Theeuwes, J.: Abrupt luminance change pops out; abrupt color change does not. *Perception & Psychophysics* 57(5), 637–644 (1995)
- [22] Treisman, A.: Features and objects: The fourteenth Bartlett memorial lecture. *Q. J. Experimental Psychology A* 40, 201–237
- [23] Turrato, M., Galfano, G.: Color, form and luminance capture attention in visual search. *Vision Research* 40, 1639–1643 (2000)
- [24] <http://www.tobii.se>
- [25] <http://www.yoyogames.com/gamemaker>