Simulation and Formal Analysis of Visual Attention in Cognitive Systems*

Tibor Bosse[1], Peter-Paul van Maanen[1,2], and Jan Treur[1]

[1] Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
peter-paul.vanmaanen@tno.nl, {tbosse, treur}@cs.vu.nl
[2] TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, the Netherlands

Abstract. In this paper a simulation model for visual attention is discussed and formally analysed. The model is part of the design of a cognitive system which comprises an agent that supports a naval officer in its task to compile a tactical picture of the situation in the field. A case study is described in which the model is used to simulate a human subject's attention. The formal analysis is based on temporal relational specifications for attentional states and for different stages of attentional processes. The model has been automatically verified against these specifications.

1 Introduction

The model of visual attention discussed and formally analysed in this paper is part of the design of a cognitive system which comprises an agent that supports a naval officer in its task to compile a tactical picture of the situation in the field. In this domain, the complex and dynamic nature of the environment makes that the officer has to deal with a large number of tasks in parallel. Therefore, in practice (s)he is often supported by agents that take over part of these tasks. However, a problem is how to determine an appropriate work division: due to the rapidly changing environment, such a work division cannot be fixed beforehand [2]. This results in a need for reallocation of work which is determined dynamically and at runtime. For this purpose, two approaches exist, namely *human-triggered* and *system-triggered* dynamic task allocation [8]. In the former case, the user can decide up to what level the system (or agent) should assist him. But especially in alarming situations the user does not have enough time to think about task reallocations [18]. In these situations it would be better if the system determines this. Hence a system-triggered dynamic task allocation is desirable.

In order to obtain such a system-triggered dynamic task allocation, the model of visual attention discussed and formally analysed in this paper can be incorporated within the supporting agent. The idea is to use an estimation of the user's current

* Parts of this paper are based on work presented at the 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'06) [5] and the Second European Cognitive Science Conference (EuroCogSci'07) [6].

L. Paletta and E. Rome (Eds.): WAPCV 2007, LNAI 4840, pp. 463–480, 2007.

attention to determine which subtasks the agent is best to pay attention to. For instance, if the user has the subtask to pay attention to a certain track on the screen, it is a possibility is that no additional support for that track is needed. In this case the agent should rather direct its own 'attention' to the user's unattended tracks. The assumption made here, that the allocation of attention actually means committing oneself to something, enables the agent to adjust its support at runtime, based on the dynamics of the modelled attention. This is a reasonable assumption, since attention is a prerequisite for conscious action [1].

It is demonstrated how such a model is used to run a simulation. This simulation is based on data from a case study in which a user executed a task abstracted from a naval radar track identification task. The present gathered data, which is only used for demonstration purposes, consist of two types of information: dynamics of tracks on a radar scope and of the user's gaze. Based on this information, the cognitive model estimates the distribution of attention levels over locations of the radar scope. Furthermore, based on the characteristics of these attention levels over time, temporal properties are defined that indicate certain attentional subprocesses,, inspired by the phases of information processing, cf. [25, 30, 31], juxtaposed to an assumption often made in literature, e.g., in [19, 34], that attention is a single, homogeneous concept.

Section 2 presents a brief introduction of the existing literature on visual attention, which helps to understand the choices made within this paper. Next, Section 3 describes the cognitive model. In Section 4, the model is illustrated by a description of a case study and the corresponding simulation results. Section 5 shows how the model can be further analysed by verifying formal temporal relational specifications for attentional states and subprocesses. Finally, Section 6 is a discussion.

2 Visual Attention

Visual attention has been a subject of study in many disciplines and this section is not intended to deliberate on all of these disciplines. It rather discusses a small but dominant part of the literature on attention, in order to bridge between relevant theory on the one hand and the application mentioned in the introduction on the other hand.

In psychology, a dominant view on attention distinguishes two types of attention: *exogenous attention* and *endogenous attention* [34]. The former stands for attention by means of triggers by (partially) unexpected inputs from the environment, i.e. bottom-up triggers, such as a fierce blow on a horn. The latter stands for attention by means of a slower trigger from within the subject, i.e. top-down triggers, such as searching a friend in a crowd. There are reasons to say that exogenous and endogenous attention are closely intertwined. A recent study [31], for instance, shows that capture of exogenous attention occurs only if the object that attracts attention has a property that a person is using to find a target.

Another relevant aspect of visual attention is the effect of so-called *inattentional blindness* [28]. This is the property that perception does not always result in attending to the important and unexpected events. Attention may also be a result of certain non-visual cognitive activities, such as having deep thoughts on history or future events. Because of the limited amount of attentional resources, this results in a blind spot for visual stimuli.

A third important discussion in the literature addresses the distinction of two definitions of visual attention: that as a division over space and that as a division over objects. The first definition is more traditional and involves continuous locations over 2D or 3D space. There are several space-based theories of attention, such as the *filter theory* [7], *spotlight theory* [32], and the *zoom-lens theory* [15], etc. They all have in common that attention is subject to whatever is within a certain location in space. The object-based view of attention is more 'recent' and stresses that attention is allocated to (groups of) perceptual objects, rather than a continuous space [14]. These objects can have various properties, such as shape, speed, colour, etc., and location is just treated as a special property of objects.

A fourth important discussion in psychology is sometimes called to be related to the *what-where-distinction* [27], and combines in some way the space- and object-based views of attention. What-attention prepares a person that something will happen concerning a certain already visible object. On the other hand, where-attention prepares the sensory memory for further deliberation. This preparation happens when a person expects something to happen in a specific region in the search space or sensor, but does not know what exactly may or will happen.

In Computer Science and Artificial Intelligence there has been a growing interest for the development and usage of mathematical models of visual attention [19]. Such models are for instance used for enhancing encryption techniques in JPEG and MPEG standards [11]. Another application is to use them for making believable virtual humans in synthetic environments [24]. Basically one can distinguish two types of questions addressed within literature on visual attention modelling:

- Given certain circumstances and behaviour, to which attention levels does this lead? Models addressing this question are for instance interesting for predicting on what aspects in a picture somebody will pay attention.
- Given certain attention levels, to which behaviour do these lead (output)? Models addressing this question are for instance interesting for generating realistic behaviour for virtual characters.

Answers to both of the above questions help in how to construct a cognitive model of visual attention. To construct such a model, several types of information may be used as input. In general, the following three types of information are distinguished:

- *Behavioural cues from the user.* The idea is that behaviour is triggered by certain attentional states. Examples of behavioural cues are gaze-duration, -frequency, -path, headpose, and task performance.
- *Properties of objects in the environment.* In that case, certain stimuli from the environment will or will not cause humans to attend to something. Examples of such cues are features of objects, such as shape, texture, colour, size, movement, direction, and centeredness. Note that this case addresses exogenous attention.
- *Properties of the human attention mechanism.* Examples of this are that humans pay attention to a speaker if they expect or want him or her to speak, or have a certain other commitment, goal, or desire. The goal is to estimate what kind of commitments, interests, goals, etc., the human has and estimate what one might expect in terms of attention levels. Note that this case addresses endogenous attention.

Next section will demonstrate how the above types of information can be integrated into one executable model.

3 A Mathematical Model for Visual Attention

In this section the mathematical model for visual attention is presented. The proposed model is composed of formal rules that are related to the psychological concepts discussed in the previous section. In Section 3.1 a formal definition of attention is given, taking into account the distinction between the two possible informal definitions stated earlier. In Section 3.2 it is described how behavioural cues of the user are derived from gaze characteristics and are used to estimate attention. In Section 3.3 saliency maps are discussed shortly, that translate properties of objects in the environment to a probable attention demand. Saliency maps are not only related to exogenous but also to endogenous attention, since saliency is task related as well. Inattentional blindness is modelled by means of fixing a certain limited amount of total attention, which is managed by normalisation, persistency, decay, and concentration processes, described in Sections 3.4, 3.5, and 3.6, respectively.

3.1 Attention Values, Objects and Spaces

As described in Section 2, there is a distinction between the definition of attention as a division over space and that as a division over objects. In this paper the first approach is used and it is assumed that one can have attention for multiple spaces at the same time. One of the reasons for using spaces instead of objects is that it is actually possible to pay attention to certain spaces that do not contain any objects (yet).

The model presented in this paper will define different (discrete) spaces, which each have a specific 'quantity' of attention. One argument for this choice is that certain spaces can contain more relevant information than others. This quantity of attention will be called the *attention value*. Division of attention is now defined as an instantiation of attention values AV for all attention spaces s. An *attentional state* is a division of attention at a certain moment in time. Mathematically, given the above, the following is expected to hold:

$$A(t) = \sum_{spaces\ s} AV(s,t)$$

where A(t) is the total amount of attention at a certain time t and AV(s,t) is the attention value for attention space s at time t. In this study we define attention spaces to be 1×1 squares within an $M \times N$ grid. In principle it holds that the more attention spaces, the less attention value for each of those spaces. This is reasonable because there is a certain upper limit of total amount of working memory humans have. In the following sections the concept of attention value is further formalised.

3.2 Gaze

As discussed earlier, human behaviour can be used to draw conclusions on a person's current attentional state. An important aspect of the visual attentional state is human

gaze behaviour. The gaze dynamics (saccades) are not random, but say something about what spaces have been attended to [9, 26]. Since people often pay more attention to the centre than to the periphery of their visual space, 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 as follows:

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

where $AV_{pot}(s,t)$ is the potential attention value of s at time point t. For now, the reader is advised to assume that $AV_{new}(s,t) = AV(s,t)$. 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):

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

Other ways for calculating attention degradation as a function of distance is for instance using a Gaussian approximation.

3.3 Saliency Maps

Still unspecified is how the potential attention value $AV_{pot}(s,t)$ is to be calculated. The main idea here is to use the properties of the space (i.e., of the types of objects present) at that time. These properties can be for instance features such as colour, intensity, and orientation contrast, amount of movement (movement is relatively well visible in the periphery), etc. For each of such a feature a specific *saliency map* describes its potency of drawing attention [11, 19, 20]. Because not all features are equally highlighting, an additional weight for every map is used. Formally the above can be depicted as:

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

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.

3.4 Normalisation

The total amount of human attention is assumed to be limited. Therefore the attention value for each space s is limited due to the attention values of other attention spaces. This can be written down as follows:

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

where $AV_{norm}(s,t)$ is called the normalised attention value for space s at time point t.

3.5 Persistency and Decay

On the one hand, visual attention is something that persists over time. If one has a look at a certain space at a certain time, it is probably not the case that the attention value of that space is lowered drastically the next moment [34]. This can be done by persistently keeping the model fed with input from the environment or the user, such as saliency and gaze, respectively. But, and this holds especially for gaze, the input is not persistent. Gaze is in general more dynamic than attention. Consider the following: reading this long sentence does not cause you to just pay attention to, and therefore comprehend, merely the characters you read, but instead, while your gaze follows specific positions in this sentence, you pay attention to whole parts of this sentence.

As a final observation, in reality it is impossible to keep one's attention to everything that one sees. In fact, given the above formulas, this will lead to increasingly low attention values (consider the formula in the previous section again).

Based on the above considerations a persistency and decay factor has been added to the model, which allows attention values to persist over time independently of the persistency of the input, but not completely: with a certain decay. Formally this can be described as follows:

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

where λ is the decay parameter that results in the decay of the attention value of s at time point t – 1. Note that higher values for λ results in a higher persistency and lower decay and vice versa.

3.6 Concentration

In this document concentration is seen as the total amount of attention one can have. For instance if for all t, A(t) = 1, then the concentration is always the same, i.e., 1. But there may be a variance in concentration. Distractions by irrelevant stimuli can be the reason for that, or becoming tired. If the model needs to describe attention dynamics precisely and the task is sensitive for irrelevant distraction, one might consider non-fixed A(t) values.

4 Case Study

Now that the model of visual attention has been explained, in this section a case study is briefly set out. The case study involves a human operator executing a naval officer-like task. For this case study, it is first explained how the data were obtained (Section 4.1). The data were then used as input for the simulation model (implemented in Matlab [16]), which is described in detail in Section 4.2. In Section 4.3 the results of the case study are shown.

4.1 Task

The model of visual attention presented above was used in a simulation run based on 'real' data from a human participant executing a naval officer-like task. The software

Multitask [12] was altered in order to have it output the proper data as input for the model. This study did not yet deal with altering levels of automation (subject of Clamann et al.'s), and the software environment was momentarily only used for providing relevant data. Multitask was originally meant to be a low fidelity air traffic control (ATC) simulation. In this study it is considered to be an abstraction of the cognitive tasks concerning the compilation of the tactical picture. A snapshot of the task is shown in Figure 1.

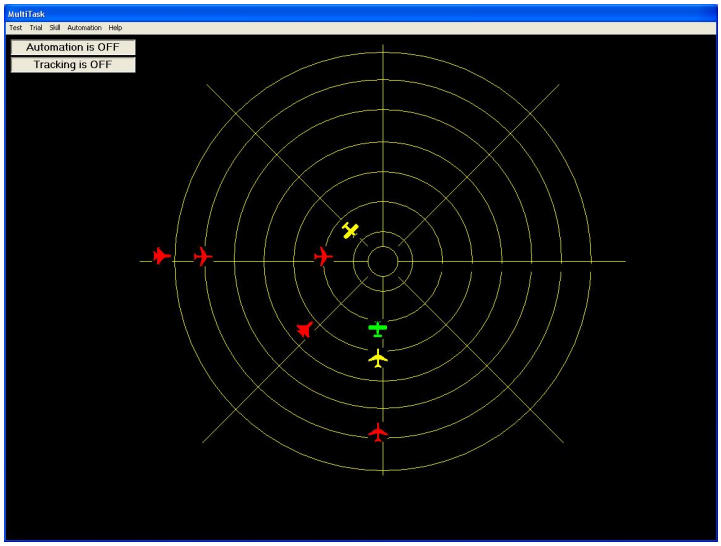

Fig. 1. The interface of the experimental environment [12]

In the case study the participant (controller) had to manage an airspace by identifying aircrafts that all are approaching the centre of a radarscope. The centre contains a high value unit (HVU) and had to be protected. In order to do this, airplanes needed to be cleared and identified to be either hostile or friendly. Clearing contained six phases: 1) a red colour indicated that the identity of the aircraft was still unknown, 2) flashing red indicated that the naval officer was establishing a connection link, 3) yellow indicated that the connection was established, 4) flashing yellow indicated that the aircraft was being cleared, 5) green indicated that either the aircraft was attacked when hostile or left alone when friendly or neutral, and finally 6) the target is removed from the radarscope when it reaches the centre. Each phase consisted of a certain amount of time and to go from phase 1 to 2 and from phase 3 to 4 required the participant to click on the left and the right mouse button, respectively. Three different aircraft types were used: military, commercial, and private. Note here that the type did not determine anything about the hostility. The different types merely resulted in different intervals of speed of the aircrafts. All of the above were environmental stimuli that resulted in change of the participant's attention.

The data that were collected consisted of all locations, distances from the centre, speeds, status of the aircrafts (which phase), and types. Additionally, data from a

Tobii x50 eye-tracker [17] were extracted while the participant was executing the task. All data were retrieved several times per second. Together with the data from the experimental environment they were used as input for the simulation model described below.

4.2 Simulation Model

To obtain a simulation model, the mathematical model as shown in Section 3 has been implemented in Matlab [16]. The behaviour of the model can be summarised as follows. Every time step (of 100 msec), the following three steps are performed:

1. First, per location, the "current" attention level is calculated. The current attention level is the weighted sum of the values of the (possibly empty) tracks on that location, divided by 1 + α * the square of the distance between the attended location and the location of the gaze, according to the formula presented in Section 3.2.
2. Then, the attention level per location is normalised by multiplying the current attention level with the total amount of attention that the person can have and dividing this by the sum of the attention levels of all locations (also see Section 3.4).
3. Finally, per location, the "real" attention level is calculated by taking into account the history of the attention. Here a constant d is used that indicates the decay, i.e., the impact of the history on the new attention level (compared to the impact of the current attention level), also see Section 3.5.

Moreover, in the simulations discussed below, the following parameter settings are used for the formulae as introduced in Section 3:

total duration of the simulation in time steps	500	
highest x-coordinate	31	
highest y-coordinate	28	
w_M(stat, t), weight factor of attribute *status* at time point t	0.8	(for all t)
w_M(dist, t), weight factor of attribute *distance* at time point t	0.5	(for all t)
w_M(type, t), weight factor of attribute *type* at time point t	0.1	(for all t)
w_M(spd, t), weight factor of attribute *speed* at time point t	0.5	(for all t)
concentration A(t), i.e. total amount of attention a person has at time point t	100	(for all t)
impact α of gaze on the current attention level	0.3	
decay parameter d, i.e., impact of history on the new attention level	0.8	

4.3 Simulation Results

The results of applying the attention model to the input data described above are in the form of an animation, see [16]. A screenshot of this animation for one selected time point (i.e., time point 193) is shown in Figure 2 (see [16] for a full colour version). This figure indicates the distribution of attention over the grid at time point 193 (i.e., 19300 msec after the start of the task). The x- and y-axis denote the x- and y-coordinates of the grid, and the z-axis denotes the level of attention. As described earlier, the grid (which originally consists of 11760x10380 pixels) has been divided in a limited (31x28) number of locations. Besides the value at the z-axis, the colour of the grid also denotes the level of attention: blue locations indicate that the location does not attract much attention, whereas green and (especially) red indicate that the location attracts more attention (see also the colour bar at the right). In addition, the

locations of all tracks are indicated in the figure by means of small "•" symbols. The colours of these symbols correspond to the colours of the tracks in the original task (i.e., red, yellow or green). Furthermore, the location of the gaze is indicated by a big blue "*" symbol, and a mouse click is indicated by a big black "●" symbol. Figure 2 clearly shows that at time point 193 there are two peaks of attention: at locations (12,10) and (16,9). Moreover, a mouse click is performed at location (16,9), and the gaze of the subject is also directed towards that location.

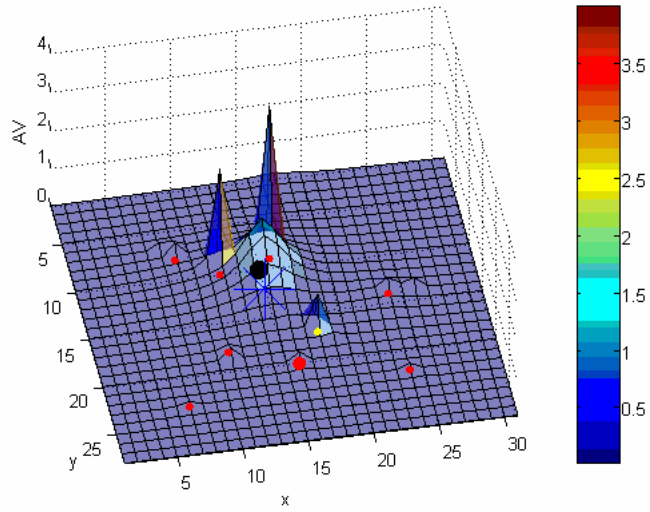

Fig. 2. Attention distribution at time point 193

5 Temporal Relational Specification and Verification

This section addresses formal analysis of the behaviour of the simulation model. To this end, it is shown how (temporal) properties of states and processes concerning visual attention can be formally specified and verified. In particular, in Section 5.1, backward and forward temporal relational specifications for attentional states are discussed, and in Section 5.2 temporal relational specifications for different attentional subprocesses. Finally, in Section 5.3 it is shown how these formally specified temporal relations can be automatically verified.

5.1 Temporal Relational Specification of Attentional States

Although the work reported in this paper focuses on a practical application context, also a formal analysis for the notion of attentional state is discussed. For this analysis the *relational specification* approach from Philosophy of Mind is adopted. This approach indicates how the occurrence of a mental state property relates to properties of states 'distant in space and time'; cf. [23, pp. 200-202]. For a relational specification for a mental state property p, two possibilities are considered:

(1) relating the occurrence of p to events in the past (*backward temporal relation*)
(2) relating the occurrence of p to behaviour in the future (*forward temporal relation*)

Applied to the case of an attentional state, a backward temporal relational specification can be used to describe what brings about this state, for example, gaze direction and cues of objects that are observed; this corresponds to (1) above. A forward temporal relational specification for attentional states, describes what the effect of this state is in terms of behaviour; this corresponds to (2) above. Below it is shown how some of these different approaches can be applied to attentional states.

A quantitative approach to mental states allows us to consider certain levels of a mental state property p; in this case a mental state property is involved that is parameterised by a number: it has the form p(a), where a is a number, denoting that p has level a (e.g., in the case considered, the amount a of attention for space s). By decay, levels decrease over time. For example, if d is the decay rate (with $0<d<1$), then at a next time point the remaining level may be d*a, unless a new contribution is to be added to the level. Decisions for certain behaviour may be based on a number of such state properties with different levels, taking into account their values; e.g., by determining the highest level of them, or the ones above a certain threshold (which may depend on the distribution of values over the different mental state properties, in the case considered here the attention levels for the different spaces).

For the *backward case*, the temporal relational specification involves a summation over different time points. Moreover, a decay rate d with $0 < d < 1$ is used. The *backward temporal relational specification* is expressed by:

> There is an amount w of attention at space s, if and only if there is a history such that at time point 0 there was initatt(0, s) attention at s, and for each time point k from 0 to t an amount new attention newatt(k, s) is added for s, and $w = \text{initatt}(0, s) * d^t + \sum_{k=0}^{t} \text{newatt}(t-k, s) * d^k$. There is an amount v of new attention for space s at t if and only if at time t-e the value v is the weighted sum of feature values for s divided by 1 plus the square of the distance of s to the gaze point and normalised for the set of spaces.

The *forward case* involves a behavioural choice that depends on the relative levels of the multiple mental state properties. This makes that at each choice point the temporal relational specification of the level of one mental state property is not independent of the level of the other mental state properties involved at the same choice point. Therefore it is only possible to provide a temporal relational specification for the combined mental state property. For the case considered, this means that it is not possible to consider only one space and the attention level for that space, but that the whole distribution of attention over all spaces has to be taken into account. The *forward temporal relational specification* is expressed as follows:

> If at time t1 the amount of attention at space s is above threshold h, then action is undertaken for s at some time $t2 \geq t1$ with $t2 \leq t1+e$ &
> If at some time t2 an action is undertaken for space s for track 1, then at some time t1 with $t2-e \leq t1 \leq t2$ the amount of attention at space s was above threshold h.

Here the threshold h can be determined, for example, as a value such that for 5% of the spaces the attention is above h and for the other spaces it is below h, or such that only three spaces exist with attention value above h and the rest under h.

5.2 Temporal Relational Specification of Attentional Subprocesses

In the previous subsection temporal relational specifications for attentional states have been defined. In recent years, an increasing amount of work is aimed at identifying *different types* of attention, and focuses not on attentional states, but on *subprocesses* of attention. For example, many researchers distinguish at least two types of attention, i.e. perceptual and decisional attention [31]. Some others even propose a larger number of functionally different subprocesses of attention [25, 30]. Following these ideas, this section provides a (temporal) differentiation of an attentional process into a number of different types of subprocesses. To differentiate the process into subprocesses, a cycle *sense – examine – decide – prepare and execute action – assess action effect* is used. It is discussed how different types of attention within these phases can be distinguished and defined by temporal specifications.

- *attention allocation*

This is a subprocess in which attention of a subject is drawn to an object by certain exogenous (stimuli from the environment) and endogenous (e.g., goals, expectations) factors, see, e.g., [34]. At the end of such an 'attention catching' process an attentional state for this object is reached in which gaze and internal focus are directed to this object. The informal temporal specification of this *attention allocation process* is as follows:

From time t1 to t2 attention has been allocated object O iff
at t1 a combination of external and internal triggers related to object O occurs,
and at t2 the mind focus and gaze are just directed to object O.

Note that in this paper validation only takes place with respect to gaze and not to mind focus, as the empirical data used have no reference to internal states.

- *examinational attention*

Within this subprocess, attention is shared between or divided over a number of different objects. Attention allocation is switched between these objects, for example, visible in the changing gaze. The informal temporal specification of this *examinational attentional process* is as follows:

During the time interval from t1 to t2 examinational attention occurs iff
from t1 to t2 for a number of different objects attention is allocated alternatively to these objects.

- *decision making attention*

A next subprocess distinguished is one in which a decision is made on which object to select for an action on a certain object to be undertaken. Such a *decision making attentional process* may have a more inner-directed or introspective character, as the subject is concentrating on an internal mental process to reach a decision. Temporal specification of this attentional subprocess involves a criterion for the decision, which is based on the relevance of the choice made; it is informally defined as follows:

During the time interval from t1 to t2 decision making attention occurs iff
at t2 attention is allocated to an object, from which the relevance is higher than a certain threshold.

- *action preparation and execution attention*

Once a decision has been made for an action, an *action preparation and execution attentional process* occurs in which the subject concentrates on the object, but this

time on the aspects relevant for action execution. The informal temporal specification is as follows:

During the time interval from t1 to t2 attention on action preparation and execution occurs iff
from t1 to t2 the mind focus and gaze is on an object O and at t2 an action a is performed for this object O.

- *action assessment attention*

Finally, after an action has been executed, a retrospective *action assessment attentional process* occurs in which the subject evaluates the outcome of the action. Here the subject focuses on aspects related to goal and effect of the action. The informal temporal specification of this attentional process is as follows:

During the time interval from t1 to t2 action assessment attention occurs iff
at t1 an action a is performed for this an object O
and from t1 to t2 the mind focus and gaze is on this object O and from t2 they are not on O.

5.3 Formal Specification and Analysis

The results of the simulation model can be analysed in detail by converting them into formally specified *traces* (i.e., sequences of events over time), and checking relevant properties, expressed in the form of the temporal relational specifications discussed above, against these traces. These properties were logically formalised in the language TTL [4]. This predicate logical language supports formal specification and analysis of dynamic properties, covering both qualitative and quantitative aspects. TTL is built on atoms referring to states, time points and traces. Dynamic properties can be formulated in a formal manner in a sorted first-order predicate logic, using quantifiers over time and traces and the usual first-order logical connectives such as $\neg$, $\wedge$, $\vee$, $\Rightarrow$ $\forall$, $\exists$. A special software environment has been developed for TTL, featuring both a Property Editor for building and editing TTL properties and a Checking Tool that enables formal verification of such properties against a set of (simulated or empirical) traces. An example of a relevant dynamic property expressed in TTL is the following:

GP1 (Mouse Click implies High Attention Level Area)
For all time points t, if a mouse click is performed at location {x,y}, then at e time points before t, within a range of 2 locations from {x,y}, there was a location with an attention level that was at least h. Here, h is a certain threshold that can be determined as explained in the previous section. Formalisation:

$\forall$t:T $\forall$x,y:COORDINATE
[state(γ,t) |= mouse_click(x,y) $\Rightarrow$ high_attention_level_nearby(γ, t-e, x, y)]

Here, high_attention_level_nearby is an abbreviation, which is defined as follows:
high_attention_level_nearby(γ:TRACE, t:T, x,y:COORDINATE) $\equiv$
$\exists$p,q: COORDINATE, $\exists$i:REAL state(γ,t) |= has_attention_level(p,q,i) &
x-2 $\leq$ p $\leq$ x+2 & y-2 $\leq$ q $\leq$ y+2 & i > h

Note that this property is a refinement of the forward temporal relational specification defined in Section 5.1. Roughly spoken, it states that for every location that the user clicks on, some time before (e time points) he had a certain level of attention. The decision to allow a certain error (see GP1: instead of demanding that there was a high attention level at the exact location of the mouse click, this is also allowed at a nearby

location within the surrounding area) was made in order to handle noise in the data. Usually, the precise coordinates of the mouse clicks do not correspond exactly to the coordinates of the tracks and the gaze data. This is due to two reasons:

(1) a certain degree of inaccuracy of the eye tracker, and
(2) people often do not click exactly on the centre of a track.

The approach used is able to deal with such imprecision.

Using the TTL Checking Tool, property GP1 has been automatically verified against the traces that resulted from the case study. For these checks, e was set to 5 (i.e. 500 msec, which by experimentation turned out a reasonable reaction time for the current task), and h was set to 0.3 (which was chosen according to the 5%-criterion, see Section 5.1). Under these parameter settings, all checks turned out to succeed. Although this is no exhaustive verification, this is an encouraging result: it shows that the subject always clicks on locations for which the model predicted a high attention level.

Besides GP1, also the temporal relations for attentional subprocesses introduced in Section 5.2 have been formalised, as shown below. To this end, first some useful help-predicates are defined:

gaze_near_track(γ:TRACE, c:TRACK, t1:TIME) ≡
  ∃x1,y1,x2,y2:COORD
    state(γ, t1) ⊨ gaze(x1, y1) &
    state(γ, t1) ⊨ is_at_location(c, x2, y2) &
    |x2-x1| ≤ 1 & |y2-y1| ≤ 1

mouseclick_near_track(γ:TRACE, c:TRACK, t1:TIME) ≡
  ∃x1,y1,x2,y2:COORD
    state(γ, t1) ⊨ mouse_click(x1, y1) &
    state(γ, t1) ⊨ is_at_location(c, x2, y2) &
    |x2-x1| ≤ 1 & |y2-y1| ≤ 1

action_execution(γ:TRACE, c:TRACK, t2:TIME) ≡
    mouseclick_near_track(γ, c, t2) &
    ∃t1:TIME t1 < t2 & ∀t3:TIME [t1 ≤ t3 ≤ t2 ⇒ gaze_near_track(γ, c, t3)]

The reason for using gaze_near_track instead of something like gaze_at_track is that a certain error is allowed in order to handle noise in retrieved empirical data. Usually, the precise coordinates of the mouse clicks do not correspond exactly to the coordinates of the tracks and the gaze data. This is due to two reasons: 1) a certain degree of inaccuracy of the eye tracker, and 2) the fact that people often do not click exactly on the, for instance, centre of a track.

Based on these intermediate predicates, the five types of attentional (sub)processes as described earlier are presented below, both in semi-formal and in formal (TTL) notation:

GP2A (Allocation of attention)
From time t1 to t2 attention has been allocated to track c iff
at t2 the gaze is directed to track c and between t1 and t2 the gaze has not been directed to any track.
has_attention_allocated_during(γ:TRACE, c:TRACK, t1, t2:TIME) ≡

  t1 < t2 & gaze_near_track(γ, c, t2) &
  ∀t3:TIME, c1 :TRACK
    [t1 ≤ t3 < t2 ⇒ ¬ gaze_near_track(γ, c1, t3)]

GP2B (Examinational attention)
During the time interval from t1 to t2 examinational attention occurs iff at least two different tracks c1 and c2 exist to which attention is allocated during the interval from t1 to t2 (resp. between t3 and t4 and between t5 and t6).

```
has_examinational_attention_during(γ:TRACE, t1, t2:TIME) ≡
  ∃t3,t4,t5,t6:TIME ∃c1,c2:TRACK
    t1 ≤ t3 ≤ t2 &  t1 ≤ t4 ≤ t2 & t1 ≤ t5 ≤ t2 &  t1 ≤ t6 ≤ t2 &
        c1 ≠ c2 &
        has_attention_allocated_during(γ, c1, t3, t4) &
        has_attention_allocated_during(γ, c2, t5, t6)
```

GP2C (Attention on decision making and action selection)
During the time interval from t1 to t2 decision making attention for c occurs iff from t1 to t2 attention is allocated to a track c, for which the saliency at time point t1 (based on features type, distance, colour and speed) is higher than a certain threshold th.

```
has_attention_on_action_selection_during(γ:TRACE, c:TRACK, t1, t2:TIME, th:INTEGER) ≡
   t1≤t2 & ∃p1,p2,p3,p4:VALUE ∀t3  [ t1≤t3≤t2 ⇒
     state(γ, t3) ⊨ has_type(c, p1) ∧ has_distance(c, p2) ∧
                    has_colour(c, p3) ∧ has_speed(c, p4) ]  &
     (0.1*p1+0.5*p2+0.8*p3+0.5*p4)/1.9 > th & has_attention_allocated_during(γ, c, t1, t2)
```

GP2D (Attention on action preparation and execution)
During the time interval from t1 to t2 attention on action preparation and execution for c occurs iff from some t4 to t1 attention on decision making and action selection for c occurred and from some t3 to t2 attention on the execution of an action on c occurs.

```
has_attention_on_action_prep_and_execution_during(γ:TRACE,
c:TRACK, t1, t2:TIME, th:INTEGER) ≡
   t1≤ t2  &  ∃t3:TIME [ t3 ≤ t1 &
     has_attention_on_action_selection_during (γ, c, t3, t1, th) ] &
     ∀t4:TIME [t1 ≤ t4 ≤ t2 ⇒ gaze_near_track(γ, c, t4)  &
     action_execution(γ, c, t2)
```

GP2E (Attention on action assessment)
During the time interval from t1 to t2 action assessment attention for c occurs iff at t1 an action on c has been performed and from t1 to t2 the gaze is on c and at t2 the gaze is not at c anymore.

```
has_attention_on_action_assessment_during(γ:TRACE, c:TRACK, t1,t2:TIME) ≡
     [ t1 ≤ t2 &
     action_execution(γ, c, t1) &
     ¬ gaze_near_track(γ, c, t2) &
     ∀t3:TIME [t1 ≤ t3 < t2 ⇒ gaze_near_track(γ, c, t3) ]
```

All the above TTL properties can be checked in the TTL Checking Tool. An example of how one could check such a property for certain parameters is the following:

```
check_action_selection ≡
    ∀γ:TRACES
      ∃t1,t2:TIME ∃c:TRACK
      has_attention_on_action_selection_during (γ, c, t1, t2, 5)
```

This property states that the phase of decision making and action selection holds for track c, from time point t1 to time point t2, with a threshold of 5, for all loaded traces. This property either holds or does not. If so, the first instantiation of satisfying parameters are retrieved.

All of the formalised dynamic properties shown above, as well as a number of additional ones (not shown due to space limitations) have been successfully checked, given reasonable parameter instantiations, against the traces. As mentioned above, although these checks cannot be seen as an exhaustive validation, they contribute to a detailed formal analysis of the simulation model. The main contribution of such an analysis is that it allows the user to distinguish different attentional states and subprocesses, which can be compared with the expected behaviour.

6 Discussion

This paper presents a cognitive model as a component of a socially intelligent agent; cf. [13]. The component allows the agent to adapt to the need for support of a naval officer for his task to compile a tactical picture. Given two types of input, i.e., user- and context-input, the implemented cognitive model is able to estimate the visual attention levels within a 2D-space. The user-input was retrieved by an eye-tracker, and the context-input by means of the output of the software for a naval radar track identification task. The first consists of the (x, y)-coordinates of the gaze of the user over time. The latter consists of the variables speed, distance to the centre, type of plane, and status of the plane. In a case study, the model was used to predict the attention of a human participant that executes the task mentioned above. The model was specifically tailored to domain-dependent properties retrieved from a task environment; nevertheless the method presented remains generic enough to be easily applied to other domains and task environments.

Although the work reported in this paper focuses on a practical application context, as a main contribution, also a formal analysis was given for attentional states and processes. To describe mental states of agents in general, the concept of representational content is often applied, as described in the literature on Cognitive Science and Philosophy of Mind; e.g., [3, 21, 22], [23, pp. 191-193, 200-202][1]. In this paper this perspective first was applied to attentional states. The general idea is that the occurrence of the internal (mental) state property p at a specific point in time is related (by a *representation relation*) to the occurrence of other state properties, at the same or at different time points. Such a representation relation, when formally specified, describes in a precise and logically founded manner how the internal state property p relates to events in the past and future of the agent. To define a representation relation, the *causal-correlational approach* is often discussed in the literature in Philosophy of Mind. For example, the presence of a horse in the field has a causal relation to the occurrence of the mental state property representing this horse. This approach has some limitations; cf. [21, 23]. Two approaches that are considered to be more generally applicable are the *interactivist approach* [3, 22] and the

[1] A more exhaustive discussion of this theme from the philosophical perspective is beyond the scope of this paper.

relational specification approach [23]. In this paper the latter approach was adopted and formalised, as it provides the flexibility and expressivity that is required to address issues as discussed below.

Fundamental issues that were encountered in the context of this work are (1) how to handle decay of a mental state property, (2) how to handle reference to a history of inputs, and (3) how to handle a behavioural choice that depends on a number of mental state properties. To address these, levelled mental state properties were used, parameterised by numbers. Decay was modelled by a kind of interest rate. Backward temporal relational specifications for attentional states were defined based on histories of contributions to attention, taking into account the interest rate. Forward temporal relational specifications for attentional states were defined taking into account combinations of multiple parameterised mental state properties, relating to the alternatives for behavioural choices. In addition, it has been shown how the notion of temporal relational specification can also be used to define and formalise different attentional subprocesses that play a role in the sense-reason-act cycle.

The temporal relational specifications have been formalised in the predicate logical language TTL. Using the TTL Checking environment, they have been automatically verified against the traces that resulted from the case study. Under reasonable parameter settings, these checks turned out to succeed, which provides an indication that the attention model behaves as desired, and allows the user to get more insight into the dynamics of attentional processes. The approach used is able to handle imprecision in the data.

This paper focused on formal analysis; although in this formal analysis also empirical data were involved, a more systematic validation of the models put forward in the intended application context will be addressed as a next step. Future studies will address the use of the attention estimates for dynamically allocating tasks as a means for assisting naval officers. To determine (in a dynamical manner) an appropriate cooperation and work division between user and system, it has a high value for the quality of the interaction and cooperation between user and system, if the system has information about the particular attentional state or process a user is in. For example, in case the user is already allocated to some task, it may be better to leave that task for him or her, and allocate tasks to the system for which there is less or no commitment from the user (yet). A threshold can facilitate a binary decision mechanism that decides whether or not a task should be supported. Open questions are related to modelling both endogenous and exogenous triggers and their relation in one model. One important element missing is for example expectation as an endogenous trigger; cf. [10, 29]. Finally, the attention model may be improved and refined by incorporating more attributes within the saliency maps, for example based on literature such as [19, 20, 33].

Acknowledgments

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

References

[1] Baars, B.J.: A cognitive theory of consciousness. Cambridge University Press, London (1988)

[2] Bainbridge, L.: Ironies of automation. Automatica 19, 775–779 (1983)

[3] Bickhard, M.H.: Representational Content in Humans and Machines. Journal of Experimental and Theoretical Artificial Intelligence 5, 285–333 (1993)

[4] Bosse, T., Jonker, C.M., van der Meij, L., Sharpanskykh, A., Treur, J.: Specification and Verification of Dynamics in Cognitive Agent Models. In: Nishida, T., Klusch, M., Sycara, K., Yokoo, M., Liu, J., Wah, B., Cheung, W., Cheung, Y.-M. (eds.) IAT 2006. Proceedings of the Sixth International Conference on Intelligent Agent Technology, pp. 247–254. IEEE Computer Society Press, Los Alamitos (2006)

[5] Bosse, T., van Maanen, P.-P., Treur, J.: A Cognitive Model for Visual Attention and its Application. In: Nishida, T. (ed.) IAT 2006. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, pp. 255–262. IEEE Computer Society Press, Hong Kong, P.R. China (2006)

[6] Bosse, T., van Maanen, P.-P., Treur, J.: Temporal Differentiation of Attentional Processes. In: Vosniadou, S., Kayser, D. (eds.) EuroCogSci. 2007. Proceedings of the Second European Cognitive Science Conference, Delphi. Greece (in press)

[7] Broadbent, D.E.: Perception and Communication. Pergamon Press, London (1958)

[8] Campbell, G., Cannon-Bowers, J., Glenn, F., Zachary, W., Laughery, R., Klein, G.: Dynamic function allocation in the SC-21 Manning Initiative Program. Naval Air Warfare Center Training Systems Division, Orlando, SC-21/ONRS&T Manning Affordability Initiative (1997)

[9] Carpenter, R.H.S.: Movements of the Eyes, Pion, London (1988)

[10] Castelfranchi, C., Lorini, E.: Cognitive Anatomy and Functions of Expectations. In: Proc. of IJCAI 2003 Workshop on Cognitive modeling of agents and multi-agent interaction, Acapulco (2003)

[11] 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)

[12] Clamann, M.P., Wright, M.C., Kaber, D.B.: Comparison of performance effects of adaptive automation applied to various stages of human-machine system information processing. In: Proc. of the 46th Ann. Meeting of the Human Factors and Ergonomics Soc., pp. 342–346 (2002)

[13] Dautenhahn, K.: Human Cognition and Social Agent Technology. John Benjamins Publishing Company, Amsterdam (2000)

[14] Duncan, J.: Selective attention and the organization of visual information. J. Exp. Psychol. 113, 501–517 (1984)

[15] Eriksen, C.W., St. James, J.D.: Visual attention within and around the field of focal attention: a zoom lens model. Perception and psychophysics 40(4), 225–240 (1986)

[16] http://www.few.vu.nl/~pp/attention

[17] http://www.tobii.se

[18] Inagaki, T.: Adaptive automation: Sharing and trading of control 147–169 (2003)

[19] Itti, L., Koch, C.: Computational Modeling of Visual Attention. Nature Reviews Neuroscience 2(3), 194–203 (2001)

[20] 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)

[21] Jacob, P.: What Minds Can Do: Intentionality in a Non-Intentional World. Cambridge University Press, Cambridge (1997)
[22] Jonker, C.M., Treur, J.: A Temporal-Interactivist Perspective on the Dynamics of Mental States. Cognitive Systems Research Journal 4, 137–155 (2003)
[23] Kim, J.: Philosophy of Mind. Westview Press, Boulder (1996)
[24] Kim, Y., Van Velsen, M., Hill Jr., R.W.: Modeling Dynamic Perceptual Attention in Complex Virtual Environments. In: Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 266–277. Springer, Heidelberg (2005)
[25] LaBerge, D.: Attentional control: brief and prolonged. Psychological Research 66, 230–233 (2002)
[26] Land, M.F., Furneaux, S.: The knowledge base of the oculomotor system. Philos. Trans. R. Soc. Londen Ser. B 352, 1231–1239 (1997)
[27] Logan, G.D.: The CODE theory of visual attention: an integration of space-based and object-based attention. Psychol. Rev. 103, 603–649 (1996)
[28] Mack, A., Rock, I.: Inattentional Blindness: Perception without Attention. In: Wright, R.D. (ed.) visual attention, Ch. 3, pp. 55–76. MIT Press, Cambridge MA (1998)
[29] Martinho, C., Paiva, A.: Using Anticipation to Create Believable Behaviour. In: Proceedings of AAAI 2006 (2006)
[30] Parasuraman, R.: The attentive brain. MIT Press, Cambridge, MA (1998)
[31] Pashler, H., Johnson, J.C., Ruthruff, E.: Attention and Performance. Ann. Rev. Psych. 52, 629–651 (2001)
[32] Posner, M.E.: Orienting of attention. Q. J. Exp. Psychol. 32, 3–25 (1980)
[33] Sun, Y.: Hierarchical Object-Based Visual Attention for Machine Vision. Ph.D. Thesis, University of Edinburgh (2003)
[34] Theeuwes, J.: Endogenous and exogenous control of visual selection. Perception 23, 429–440 (1994)