Personalisation of Computational Models of Attention by Simulated Annealing Parameter Tuning

Rianne van Lambalgen
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
email: rm.van.lambalgen@few.vu.nl

Peter-Paul van Maanen
Vrije Universiteit Amsterdam
Amsterdam The Netherlands
and TNO Human Factors
Soesterberg, The Netherlands
email: peter-paul.van.maanen@tno.nl

Abstract -- **In this paper it is explored whether personalisation of an existing computational model of attention can increase the model's validity. Computational models of attention are for instance applied in attention allocation support systems and can benefit from this increased validity. Personalisation is done by tuning the model's parameters during a training phase, using Simulated Annealing (SA). The adapted attention model is validated using a task, varying in difficulty and attentional demand. Results show that the attention model with personalisation results in a more accurate estimation of an individual's attention as compared to the model without personalisation.**

Attention Model, Personalisation, Parameter Tuning

I. INTRODUCTION

In a critical domain such as that of Naval Warfare, it is important for the crew to be aware of the situation on the field. However, the person has to deal with a large number of tasks in parallel and often the radar contacts are simply too numerous and dynamic to be adequately monitored by a single human. In [1], a simulation-based environment is presented that is similar to this Naval domain. In this environment, a variable amount of contacts have to be monitored, identified and handled.

Since attention is typically directed to one bit of information at a time [2], [3], a supporting software agent can be used. The agent alerts the human about a contact if it is ignored. To this end the agent has to maintain a model of the cognitive state of the human including the human's distribution of attention. In [4], a validation is done on such a support agent by using a task called the Tactile Picture Compilation Task. It is shown that the designed support agent indeed improves the human's performance.

The existing model of the attention distribution (see also [5]) is static in the sense that parameter values are set beforehand. However, it is known that such parameters may depend on personal characteristics and therefore it is useful to adjust them for each person performing a task. The focus of this research is to personalise the attention model by using Simulated Annealing (SA) to tune these parameters. Earlier work on validation of a cognitive model shows that appropriate parameter values can be found using SA [6].

Personalisation of the attention model is done by tuning the model's parameters during a training phase. For validation of the adapted attention model, an experiment is conducted in the simulation-based environment. The task participants had to perform varied in difficulty and attentional demand. To obtain results on a person's attention, participants had to indicate the objects that had their attention at a certain moment in time.

In the first section of this paper, the existing attention model is shown and a theoretical background is given on individual differences in attention. In section III the task and procedure of the experiment are described, followed by a description of the applied SA algorithm in section IV. Next, section V compares the validity of the attention model with personalisation to that of the attention model without personalisation. The results are discussed in section VI.

II. ATTENTION MODEL

A. *Description of the Attention Model*

The Attention Model is taken from [4] and is briefly summarized in this section. The Attention Model is a mathematical model that uses input from features of objects on the screen and an agent's gaze to provide an estimation of the current attention distribution at a time point: an assignment of attention values AV(s, t) to a set of attention spaces (i.e. areas on a computer screen) at that time. The attention distribution is assumed to have a certain persistency. At each point in time the new attention is related to the previous attention:

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

Here, λ is the decay parameter for the decay of the attention value of space s at time point t – 1, and $AV_{norm}(s,t)$ is determined by normalisation for the total amount of attention, described by:

$$AV_{norm}(s,t) = \frac{AV_{new}(s}{\sum_{s'} AV_{new}(} \quad (2)$$

$$AV_{new}(s,t) = \frac{AV_{pot}(s,t)}{1+\alpha \cdot r(s,t)^2} \quad (3)$$

Here $AV_{new}(s, t)$ is calculated from the potential attention value of space s at time point t and the relative distance of each space s to the gaze point (the centre). 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):

$$r(s,t) = d_{eucl}(\tag{4}$$

The potential attention value $AV_{pot}(s,t)$ is based on the features of the space (i.e., of the types of objects present) at that time (e.g., luminance, colour):

$$AV_{pot}(s,t) = \sum_{maps\ M} M(s,t) \cdot w_M \tag{5}$$

For every feature there is a saliency map M, which describes its potency of drawing attention (e.g. [7]). Moreover, M(s,t) is the unweighted potential attention value of s at time point t, and $w_M(s,t)$ is the weight used for saliency map M, where $1 \leq M(s,t)$ and $0 \leq w_M(s,t) \leq 1$.

B. *Individual Differences in Attention*

Previous literature shows that individual differences in working memory capacity and general intelligence result in differences in controlled attention [8], [9]. Controlled attention (i.e. top-down) can be seen as the ability to focus attention and the ability to prevent it to be captured by other events (mental or environmental). This indicates that for individuals with high working memory capacity, features that involve top-down attention will attract more attention as compared to features that automatically capture attention (e.g. luminance, color). As a consequence, those individuals will show less switching of attention from one location to the other. In the attention model, this could indicate a lower value for the decay (λ) of attention at a location.

In addition, it is known that factors like exhaustion and experienced pressure (i.e. a human's functional state, [6]) influence the human performance and attention [10]. For example, the experienced pressure of a person can cause tunnel vision (narrowing of the attentional field), which has effect on α; the impact of the distance from the gaze point to a location on the attention value at that location. Since the functional state can differ across individuals, these factors should be taken into account in the estimation of attention.

Considering these differences, the attention model should be adjusted for each individual. In order to obtain a personalized model, in this paper parameter tuning is performed to a number of parameters described in the attention model in section II-B. As explained in this section, the delay parameter can depend on the individual and will therefore be tuned.

Furthermore, parameters are tuned that concern the weight from the saliency map of a feature to the potential attention value. Values of these weight variables may differ between individuals with low versus high memory capacity. The exact parameters that are tuned to obtain a personalised attention model, depend on the specific task and are described in the parameter tuning section.

III. METHOD

A. *Simulation-based training Environment*

The Simulation-based training Environment [1] that was used in this study consists of identifying incoming contacts and, based on the outcome of identification, deciding to eliminate the contact (by shooting) or allowing it to land (by not shooting). The participant controls a (stationary) weapon located at the bottom of a computer screen. In addition, contacts (allies and enemies in the shape of a dot) appear at a random location on the top of the screen and fall down to a random location at the bottom of the screen. A screenshot of the environment is shown in figure 1.

Before a contact can be identified, it has to be perceived. This is done by a mouse click at the contact, which reveals a mathematical equation underneath the contact. The identification task is to check the correctness of the mathematical equation (which is less difficult in less demanding situations). A correct equation means that the contact is an ally; an incorrect equation indicates that the contact is an enemy. Identification is done by pressing either the left or right arrow for respectively an ally or enemy. When a contact is identified, a green (for an ally) or a red (for an enemy) circle appears around the contact.

The contacts that have been identified as an enemy have to be shot before they land. A missile is fired by executing a mouse click at a specific location; the missile will move from the weapon to that location and explode exactly at the location of the mouse click. When a contact is within a radius of 50 pixels of the exploding missile, it is destroyed.

In this scenario participants have to pay attention to the most important contacts on the screen for an optimal performance of the task. However, when the number of contacts is large, this is not always possible. In addition, the task is also cognitively challenging, given the mathematical equations that have to be classified as either correct or incorrect. These characteristics are similar to real-life situations (e.g. air traffic control), which allows us to test the attention model in a situation close to reality.

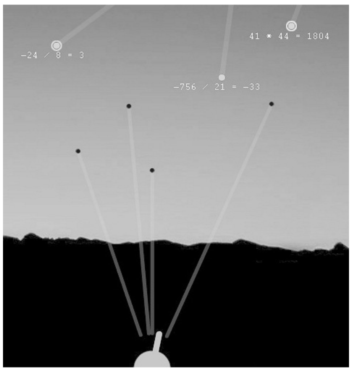

Figure 1. Screenshot of the simulation-based training environment.

TABLE I. OVERVIEW OF THE TUNED PARAMETERS

Parameter	Description
Speedweight	Weight of speed of an object (w_S in formula 5)
Distanceweight	Weight of distance of an object (w_D in formula 5)
Friendweight	Weight of the identity of an object (friendly or hostile) (w_F in formula 5)
Gazeweight	Weight of a person's gaze at a location (α in formula 3)
Decayfactor	Factor of keeping attention at a location (λ in formula 1) location
Taskfactor	Task influence when there are no objects in an attention space
Amount of ms before freeze	The time a person takes into account when asked which contacts 'recently' got their attention.

B. Participants and procedure

In this study, 2 female participants and 3 male participants with a mean age of 22.8 took part. All participants already had some experience with the experimental environment.

The experiment consisted of 2 blocks of 20 minutes of the Experimental Task. In the first 10 minutes of one block, task demands were low (contacts appear every 10 to 20 seconds) and in the second 10 minutes of one block, task demands were high (contacts appear every 2.25 to 4.5 seconds). Furthermore, in the condition with high demands, mathematical equations were more difficult than in the condition with low demands.

In both blocks, 'freezes' were made after each 2.5 minutes. When a freeze was made, the experiment was put on hold and the following sentence was shown: "Gameplay frozen. Select contacts, press space when done." At this moment, participants had to select all contacts which they thought to have recently paid attention to. After selection, a computer version of the NASA-TLX was shown, where participants had to indicate their performance and mental effort.

Throughout the experiment, eye gaze was measured using a Tobii x.60 eye tracker. For this, calibration was performed at the start of the experiment. When eye tracking was optimal, onscreen instructions were given on the task environment and freezes. The instructions were followed by a practice block of two minutes medium task demands to get familiar with the environment. After practice, participants started with the first block. After the first block, the participant was given a three minute break before continuing with the next block.

IV. PERSONALISATION OF COMPUTATIONAL MODELS OF ATTENTION

For data analysis, predictions of the two attention models (with and without personalisation) were checked and compared for validation. The parameters that are tuned in the attention model are listed in section IV-A and a description of the procedure to personalise attention models is given in Section IV-B. The evaluation procedure used to check the models for validity is described in Section IV-C, and the eventual data analyses is described in Section IV-D.

A. Parameters to be tuned

In table I, an overview is given of the parameters that are tuned for personalisation. Note that the weight values represent the weight of a feature from a saliency map to the potential attention value of that feature (formula 5 of the attention model).

B. Simulated Annealing Parameter Tuning

For personalising the attention model for each participant, Simulated Annealing was used. This method uses a probabilistic technique to find a parameter setting (see description of the parameters of the attention model in section IV-A). The parameter setting of the model without personalisation is chosen as the best available parameter setting at the start. These settings were proven to be properly set by hand in different pilots.

After the initial phase a replacement is introduced into these settings to generate a neighbour of the current parameter settings in the search space. To limit the search space, upper and lower bounds were introduced for each parameter (see table II). This leads to the following formula:

```
neighbour(x, bounds, jumpfactor) = X
+ times( randperm(length(X)) == length(X),
(bounds(:,2) - bounds(:,1))' )
* randn * permfactor                                    (6)
```

where X is a vector of the current parameter setting, times is vector multiplication, randperm(n) returns a random permutation of integers until n, bounds(:,1) and bounds(:,2) are the lower and upper bounds, randn is a random number between 0 and 1, and permfactor is the permutation speed.

TABLE II. BOUNDS OF MODEL PARAMETERS

Parameter	Upper Bound	Lower Bound
Speedweight	0	2
Distanceweight	0	2
Friendweight	0	2
Gazeweight	0	140
Decayfactor	0.8	1
Taskfactor	0	1
Amount of Ms before freeze	1500	3500

If the neighbour is found to result in a higher validity (i.e. a lower energy value E) of the model (described later in Section IV-C) then it is marked as the best known parameter setting. Otherwise the neighbour can still be selected with a small probability dependent on a decreasing temperature value. The best parameter setting is always stored, also when a different neighbour is selected. When a neighbour is not selected a new neighbour is generated to evaluate its appropriateness, and so on, until certain stopping criteria hold. Eventually the set of best parameters converges to the optimal solution (i.e. due to the probabilistic character of SA, the found solution is semi-optimal). The algorithm is described as follows:

Algorithm: SA-PARAMETER-TUNING

Input:
Initial parameter vector **X**,
maximum computation steps $\mathbf{C_{max}}$,
observed human behaviour **B**,
minimum temperature $\mathbf{T_{min}}$,
initial temperature **T**,
maximum successes $\mathbf{S_{max}}$,
maximum consecutive rejections $\mathbf{R_{max}}$,
bounds **b**, jump factor **j**;

Output:
Best estimate of parameter settings X_{best};
$\mathbf{X_{best}} = \mathbf{X}$; $\mathbf{X_{select}} = \mathbf{X}$; C = 1; **S** = 0; **E** = 1; $\mathbf{E_{best}}$ = 1; **R** = 0;

While $\mathbf{C_{max}} \geq \mathbf{C}$ and $\mathbf{T} \geq \mathbf{T_{min}}$ and $\mathbf{S_{max}} \geq \mathbf{S}$ and $\mathbf{R_{max}} \geq \mathbf{R}$ do
  While **X** not changed or not within bounds do
    $\mathbf{X_{new}}$ = neighbour($\mathbf{X_{select}}$, **b**, **j**); // see formula 6
    $\mathbf{E_{new}}$ = 1 – auc-evaluate($\mathbf{X_{new}}$, **B**); // see next section
  If $\mathbf{E_{new}} < \mathbf{E_{best}}$ do
    $X_{best} = X_{new}$; $E_{best} = E_{new}$;
  If $\mathbf{E_{new}} < \mathbf{E_{selected}}$ do
    $X_{selected} = X_{new}$; $E_{selected} = E_{new}$;
    Increase **S**;
    Reset **R**;
  Else do
    If random < $e^{(\mathbf{Eselected - Enew})/\mathbf{T}}$ do
      $X_{selected} = X_{new}$; $E_{selected} = E_{new}$;
      Increase **S**;
  Else do
    Increase **R**;
  Decrease **T**; Increase **C**;
Return $\mathbf{X_{best}}$;

This algorithm has been implemented using Matlab and the used scripts and data can be found at [11].

TABLE III. CONFUSION MATRIX

		Participant		
		t	f	total
Model	t'	Hits	False Alarms	T'
	f'	Misses	Correct Rejections	F'
	total	T	F	

C. *Subjective Evaluation Measure*

The validity was measured based on a subjective measure of attention (in terms of personal reports by the participants during the freezes). The output of the models have been compared with subjective data retrieved during the freezes in the experiment. This means that both the models as well as the participants indicated where the attention of the participant was allocated to, and these indications were compared with each other afterwards.

The comparison of the models with the subjective data was done using Area Under the Curve (AUC) analysis. AUCs are solutions of the integrals of the Receiver-Operator Characteristic (ROC) curves. ROC curve analysis was adopted from signal detection theory, where the sensitivity and specificity of detecting a signal, i.e. in our case the detection of the fact that the participant was paying attention to an object, is quantified and visualized [15]. The ROC curves can be drawn by plotting sensitivity (hit rate) versus 1 – specificity (false alarm rate). The hit rate and false alarm rate can be extracted from confusion matrices, as is shown in table III.

In this table, *t* and *f* represent whether the participant indicated that he allocated attention to an object or not, respectively, and *t’* and *f’* indicate that one of the models indicated it or not, respectively. For drawing the ROC curve, *Hits/T* results in the hit rate and *False Alarms/F* results in the false alarm rate.

The procedure of testing the hypotheses based on the subjective data was as follows:

Algorithm: AUC-EVALUATE

Input:
Parameter vector **X**,
observed human behaviour **B;**

Output:
Area under the curve **AUC**;

Construct confusion matrices (10^5) for each:
  Participant (5)
  Model type (2): with and without personalization
  Decision threshold (10^4): if **AV**(**x**,**y**) > **th**, then ***t’*** holds

Plot ROC curves (5×2 = 10, using 10^5 confusion matrices)

Calculate area under the curve (**AUC**) for each ROC curve (where 1 = good, 0 = poor)

Return the average over **AUC**s per condition (2 conditions, per condition 5 **AUC**s)

The above procedure has also been implemented using Matlab. The used scripts and data can be found at [11].

D. Data Analysis

In order to determine whether personalisation is beneficial in terms of validity, for each participant the 5 AUCs per condition for each model type is compared with each other, using a paired one-sided t-test. The model without personalisation used fixed parameters and the model with personalisation used the for each participant tuned parameters.

V. RESULTS

The personalisation procedure has been applied to all participants as mentioned in Section III-B. The result is shown in figure 2 where the energy of the best parameter settings for each participant is set out against the number of computation steps. As you can see, the results converge to the optimum (zero energy).

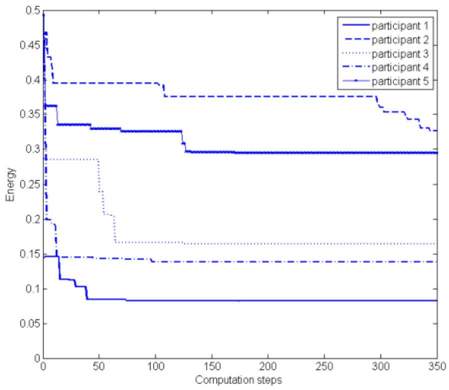

Figure 2. Best energy values per participant during SA parameter tuning

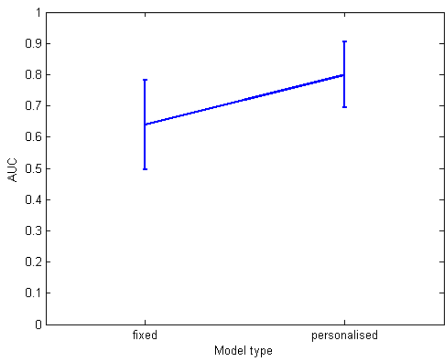

Figure 3. Means area under the curve (AUC) for the model without (left) and with (right) personalisation.

Figure 3 shows the mean AUCs for the model with and without personalisation. A paired one-sided t-test showed that the personalised attention models have a significantly mean AUC compared to the fixed attention models ($t(8) = 2.00$, $p = 0.042$). Hence, based on the used subjective validation data, the hypothesis that personalisation of attention models indeed can be accepted.

VI. DISCUSSION

In this paper, a previously presented attention model has been applied on a dynamic task, varying in attentional demands. In addition, personalisation, using Simulated Annealing parameter tuning, has been applied on the model. Validation results showed a significant difference between the personalised attention model and the fixed attention model, indicating that the former (personalised) was better in predicting attention as compared to the latter (fixed).

The validation was subjective in the sense that a participants' own estimation was measured by asking to which objects they had directed their attention before the given freezes. The subjective measure makes the assumption that people have good introspection skills, i.e., that they are accurate in the estimation of their own attention distribution. Although various experiments have pointed out that this is a reasonable assumption, it makes sense to analyse the results using an objective measure as well. A possible way of measuring objective attention is by looking at mouse clicks at a location. In future research, this can be taken into account, in addition to the subjective measure.

Simulated Annealing parameter tuning has proven to be effective in estimating a person's attention in this specific task. The same algorithm is previously used to obtain a personalised model for a Human's Functional State [6]. The application to two different models shows that the SA algorithm can be used for parameter tuning for different situations and different models. However, it should be noted that SA is a probabilistic procedure and therefore is suboptimal, specifically as the necessary computing capacity becomes relatively smaller compared to the problem space.

In the literature, personalisation is commonly used, but mainly in the area of human-computer interaction to obtain a user model for a personalised user interface [12]. Also, in previous research, neural networks are used to estimate the operator functional state [13]. However, although attention models have previously been proposed [8, 14], the authors know of no attention model that has been adjusted to the individual.

In the future personalisation of attention models can be extended. In the current personalised model parameters are tuned that are known to differ per individual. However, in future research personalisation can be done by using collected data on personality and using that to improve the attention model. Furthermore, in the current personalised model, parameters like the attention threshold and the total amount of attention are static. These could be coupled to a individual's functional state (e.g. experienced pressure, exhaustion), making the model fit for each individual, but also in different conditions (high/low workload). Such adjustments are expected to result in again an increase of the model's validity.

VII. References

[1] Both, F., Hoogendoorn, M., Lambalgen, R. van, Oorburg, R., and Vos, M. de, Relating Personality and Physiological Measurements to Task Performance Quality. In: N.A. Taatgen and H. van Rijn (eds.), Proc. of the 31th Annual Conference of the Cognitive Science Society, CogSci'09. Cognitive Science Society, Austin, TX, 2009, pp. 2819-2825.

[2] Treisman, A. Features and objects: The 14th Bartlett memorial lecture. Q.J. Experimental Psychology A. 40, 201-237.

[3] Posner, M. E., Orienting of attention, Q. J. Exp. Psychol., 32, 1980, pp. 3-25.

[4] Bosse, T., Lambalgen, R. van, Maanen, P.P. van, and Treur, J., *Attention Manipulation for Naval Tactical Picture Compilation*. In: Baeza-Yates, R., Lang, J.,

[5] Bosse, T., Maanen, P.P. van, and Treur, J., (2007). Simulation and Formal Analysis of Visual Attention in Cognitive Systems. In: L. Paletta, E. Rome (eds.), *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint, Proceedings of the Fourth International Workshop on Attention in Cognitive Systems, WAPCV'07*. Lecture Notes in Artificial Intelligence, vol. 4840. Springer Verlag, 2007, pp. 463-480.

[6] Both, F., Hoogendoorn, M., Jaffry. S.W., Lambalgen, R. van, Oorburg, R., Sharpanskykh, A., Treur, J., and Vos, M. de, Adaptation and Validation of an Agent Model of Functional State and Performance for Individuals. In: Proceedings of 12th International Conference on Principles of Practice in Multi-Agent Systems. (PRIMA'09), Nagoya, (Japan), 2009, to appear.

[7] Itti, L. and Koch, C. (2001). Computational Modeling of Visual Attention, *Nature Reviews Neuroscience*, Vol. 2, No. 3, 2001, pp. 194-203.

[8] Kane, M.J., Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychological Bulletin & Review 9(4),* 637-671.

[9] Barrett, L.F., & Tugade, M.M., & Engle, R.W. (2004). Individual differences in working memory capacity and dual-process theories of mind. *Psychological Bulletin 130(4)*, 553-573.

[10] Hockey, G.R.J. 2003. Operator Functional State as a framework for the assessment of performance degradation. In Operator Functional State. G.R.J.

[11] http://www.few.vu.nl/~pp/attention/ECAI10/

[12] Henricksen, K., & Indulska, J. Personalising context-aware applications. In: R.Meersman et. Al. (Eds.): *OTM Workshops, 2005, LNCS 3762, pp 122-131,* 2005.

[13] Wilson, G.F., & Russell, C.A. Operator Functional State classification using multiple psychophysiological features in an air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 45 (3),* 381-389 (2003).

[14] Chen, L.Q., Xie, X., Fan, X., Ma, W.Y., Zhang, H.J., and Zhou, H.Q., A visual attention model for adapting images on small displays, *ACM Multimedia Systems Journal*, 2003.