MODELING VISUAL SEARCH: EVOLVING THE SELECTIVE ATTENTION FOR IDENTIFICATION MODEL (SAIM)

DIETMAR HEINKE, GLYN W. HUMPHREYS, CLAIRE L. TWEED*

Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom,
E-mail: d.g.heinke@bham.ac.uk

We present an extension of the Selective Attention for Identification model (SAIM) [1] in which feature extraction processes are incorporated. We show that the new version successfully models experimental results from visual search. We also predict the influence of a target cue on search. This extended version of SAIM may provide a powerful framework for understanding human visual attention

1. Introduction

Recently, we have presented a connectionist model of human visual attention, termed SAIM (Selective Attention for Identification Model) [1]. SAIM's behaviour is dominated by interactions between processing units within and between modules compete to control access to stored representations for object recognition. SAIM provides a qualitative account of a range of psychological phenomena on both normal and disordered attention. Simulations on normal attention match psychological data on: two-object costs on selection, effects of object familiarity on selection, global precedence, spatial cueing both within and between objects, and inhibition of return. When simulated lesions were conducted, SAIM also demonstrated both unilateral neglect and spatial extinction, depending on the type and extent of the lesion. Different lesions also produced view-centred and object-centred neglect, capturing the finding that both forms of neglect can occur within a single patient. In essence, SAIM suggested that attentional effects in human behaviour result from competitive interactions in visual selection for object recognition, whilst neurological disorders of

*This paper was supported by grants from the BBSRC, the EPSRC and MRC, UK.

selection are due to imbalanced competition following damage to areas of the brain modulating access to stored knowledge.
In this paper we present an extended version of SAIM in which a feature extraction process was added whilst at the same time maintaining the basic principles of SAIM, (e.g. competitive interactions and selection mechanism). Our aim here is to demonstrate that this new version still successfully performs translation-invariant object identification. Additionally, we assess the viability of 'extended SAIM' as a psychological model, testing whether it can simulate and explain data from human visual search tasks.
Visual search is a commonly-used paradigm in psychological studies of attention in which participants are asked to report the absence or presence of a specified target item amongst irrelevant items (distractors). Typically, performance is measured in terms of time until response (reaction time). The number of distractors is varied across trials. A typical outcome of many experiments is a linear function between reaction time and number of distractors. The slope of this linear relation is often interpreted as an indicator of the underlying search mechanism. For instance, a small slope (0-10ms/item) is interpreted as parallel search and a steep slope (20-50ms/item) is assumed to indicate serial search, based on one item at a time (see [2], for a recent review).

2. SAIM

2.1. *Overview*

Figure 1 gives an overview of SAIM's architecture and highlights the modular structure of SAIM with each module.
In the first stage features are extracted from the input image (the feature extraction process). The contents network maps a section of the features into a smaller Focus of Attention (FOA), a process modulated by spatial attention. In addition the mapping of the contents network into the FOA is translation-invariant, enabling SAIM to perform translation-invariant object recognition. The selection network controls the contents network by competitive interactions between its processing units, so that input from only one (set of) locations is dominant and mapped into the FOA. At the top end of the model, the knowledge network identifies the contents of the FOA using template matching. The knowledge network also modulates the behaviour of the selection process with top-down activation, so that known objects are preferred over unknown objects. In addition to these modules,

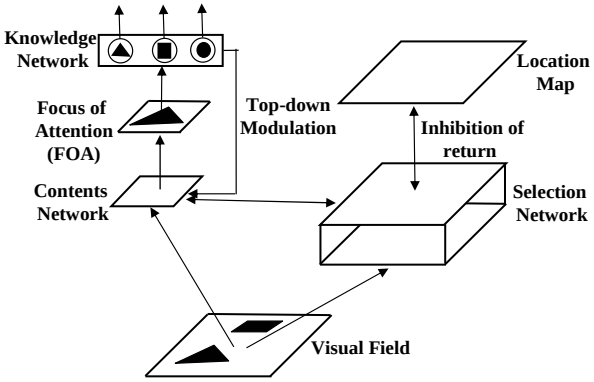

Figure 1. Architecture of SAIM

there is also a location map that enables SAIM to make multiple selections. Essentially units in the location map store the object position each time an object is recognized and then inhibits the selection network from reselecting these locations (inhibition of return).

The design of SAIM's network follows the idea of soft constraint satisfaction in neural networks based on "energy minimization" [3]. In SAIM the "energy minimization" approach is applied in the following way: Each module in SAIM performs a pre-defined task (e.g. the knowledge network has to identify the object in the FOA). In turn each task describes allowed states of activation in the network. These states then define the minima in an energy function. To ensure that the model as a whole satisfies each constraint, set by each network, the energy functions of each network are added together to form a global energy function for the whole system. The minima in the energy function is found via gradient descent, as proposed by [3]. In the following sections the energy functions for each network are stated. The global energy function and the gradient descent mechanism are omitted, since they are clearly defined by the subcomponents of the energy function.

2.2. *Feature extraction*

The feature extraction results in a three-dimensional feature vector: horizontal and vertical lines and the image itself. The lines are detected by filtering the image with 3x3 filters $\begin{pmatrix} -2 & +1 & -2 \\ -2 & +1 & -2 \\ -2 & +1 & -2 \end{pmatrix}$ for vertical lines and its

transposed version for horizontal lines). The feature vector is noted as f_{ij}^n hereafter, with indices i and j refereing to retinal locations and n to the feature dimension. This feature extraction process provides an approximation of simple cell responses in V1. As becomes obvious in the following sections, the use of just this simple feature extraction is not of theoretical value and arises only from practical consideration (e.g., the duration of any simulations). In principle, a more biologically realistic feature extraction process can be substituted (e.g. Gabor filter).

2.3. *Contents network*

The energy function for the contents network is:

$$E^{CN}(\mathbf{y^{SN}}, \mathbf{y^{CN}}) = \sum_{ijlm} (y_{lmn}^{CN} - f_{ij}^n)^2 \cdot y_{lmij}^{SN} \tag{1}$$

y_{lmij}^{SN} is the activation of units in the selection network and y_{lmn}^{CN} is the activation of units in the contents network. Here and in all the following equations the indices i and j refer to retinal locations and the indices l and m refer to locations in the FOA. The term $(y_{lmn}^{CN} - f_{ij}^n)^2$ ensures that the units in the contents network match the feature values in the input image. The term y_{lmij}^{SN} ensures that the contents of the FOA only reflect the region selected by the selection network ($y_{lmij}^{SN} = 1$). Additionally, since setting an arbitrary choice of y_{lmij}^{SN}s to 1 allows any location to be routed from the feature level to the FOA level, the contents network enables a translation-invariant mapping.

2.4. *Selection network*

The mapping from the retina to the FOA is mediated by the selection network. In order to achieve successful object identification, the selection network has to fulfill certain constraints when it modulates the mapping process. These constraints are that: (i) units in the FOA should receive the activity from only one retinal unit; (ii) activity of retinal units should be mapped only once into the FOA; (iii) neighbourhood relations in the retinal input should be preserved in mapping through to the FOA. Now, to incorporate the first constraint, that units in the FOA should receive the activity of only one retinal unit, the equation of the WTA-equation suggested by [4] turns into:

$$E_{WTA}^{SN1}(\mathbf{y^{SN}}) = \sum_{ij} (\sum_{lm} y_{lmij}^{SN} - 1)^2 \tag{2}$$

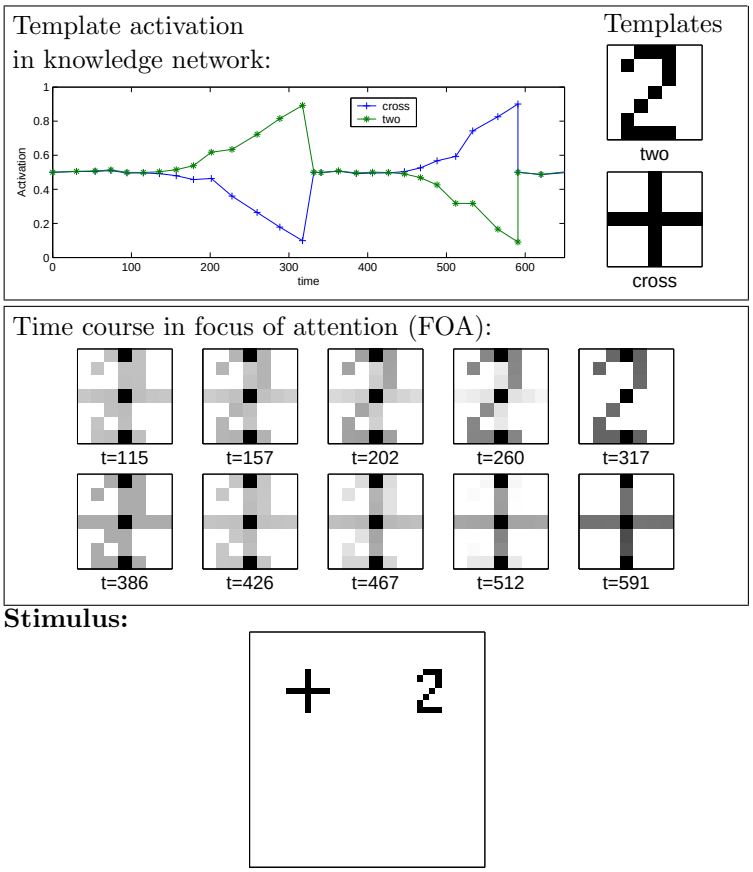

Figure 2. Basic behaviour of new SAIM.

The second term implements the second constraint:

$$E^{SN2}_{WTA}(\mathbf{y^{SN}}) = \sum_{lm}(\sum_{ij} y^{SN}_{lmij} - 1)^2 \tag{3}$$

In both terms the expression $(\sum y^{SN}_{ikjl} - 1)^2$ ensures that the activity of one location is mapped only once into the FOA.
The energy following energy function implements the neighbourhood constraint:

$$E^{SN}_{neighbor}(\mathbf{y^{SN}}) = -\sum_{i,j,l,m}\sum_{\substack{s=-L\\ s\neq 0}}^{L}\sum_{\substack{r=-L\\ r\neq 0}}^{L} g_{sr}\cdot y^{SN}_{lmij}\cdot y^{SN}_{i+r,k+s,j+r,l+s} \tag{4}$$

with g_{sr} being defined by a Gaussian function:

$$g_{sr} = \frac{1}{A} \cdot e^{-\frac{s^2+r^2}{\sigma^2}} \tag{5}$$

where A was set, so that the sum over all g_{sr} is 1. When units linked via g_{sr} are activated to $y^{SN}_{lmij} = 1$, the energy is smaller than when these units have different values, e.g. zero and one. Since g_{sr} connects units that relate to adjacent locations in both the FOA and the input image, this implements the neighbourhood constraint.
To implement inhibition of return, the location map prevents the reselection of an inhibited location through the following energy function:

$$E^{SN3}(\mathbf{y^{SN}}) = -\sum_{lmij}(1 - \sum_{lm} y^{LM}_{ij}) y^{SN}_{lmij} \tag{6}$$

The term $(1 - y^{LM}_{ij})$ suppresses already-selected locations and supports the selection of new locations.

2.5. *Knowledge network*

The energy function of the knowledge network is defined as

$$E^{KN}(\mathbf{y^{KN}}, \mathbf{y^{CN}}) = a^{KN}(\sum_{k} y^{KN}_{k} - 1)^2 - b^{KN} \sum_{lmn}(y^{CN}_{lmn} - w^{k}_{lmn})^2 y^{KN}_{k} \tag{7}$$

The index k refers to template units whose templates are stored in their weights (w^{k}_{lmn}). The term $(\sum_k y^{KN} - 1)^2$ restricts the knowledge network to activate only one template unit. The term $\sum_{lmn}(y^{CN}_{lmn} - w^{k}_{lmn})^2 \cdot y^{KN}_{k}$ ensures that the best-matching template unit is activated. a^{KN} and b^{KN} weight these constraints against each other.

2.6. *Rechecking*

In order to implement rechecking, a "location map" is computed based on activity in the selection network:

$$y^{LM}_{ij} = y^{LM}_{ij}(old) + a^{IR} \sum_{l=1}^{M} \sum_{m=1}^{M} y^{SN}_{lmij} \tag{8}$$

When a template unit in the knowledge network passes a threshold θ, the location map is used to reduce the activity in the visual field. a^{IR} controls the amount of inhibition. All units in the selection network and the knowledge network are set to the initial state they had at the beginning of the simulation.

3. Results and discussion

3.1. *Basic behaviour*

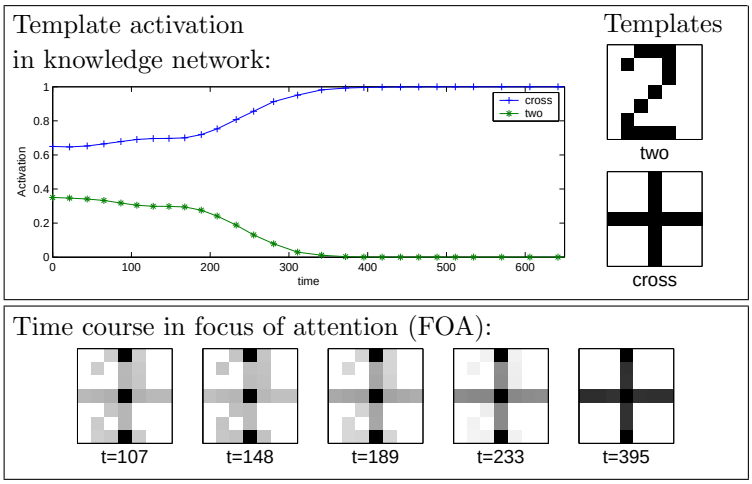

Figure 3. The simulation shows how higher initial values for the crossover comes the bottom-up bias for the two (see Fig. 2). The input image was the same as in Fig. 2

Fig. 2 demonstrates the basic behaviour of the new version of SAIM, when presented with two objects (a two and a cross). It shows that both objects are selected in a serial manner, in this case the two followed by the cross. Similarly to SAIM version 1 [1]there was a bottom-up preference towards a certain objects, here the two. This bottom-up preference results from a combination of the dynamics of the selection network and the feature extraction. We do not claim psychological plausibility for the two being preferred in selection over the cross, but it does illustrate asymmetries in bottom-up bias in the model. Fig. 3 shows that the bottom-up bias can be altered by giving the cross-template a higher initial value. This higher activation filters through the selection network via the top-down modulation (see Fig. 1). The top-down bias was used in simulations of visual search tasks where SAIM was required "to look for a target".

3.2. *Visual Search*

Fig. 4 shows that the new version of SAIM is capable of simulating typical results of visual search experiments (see [2] for examples), with linear

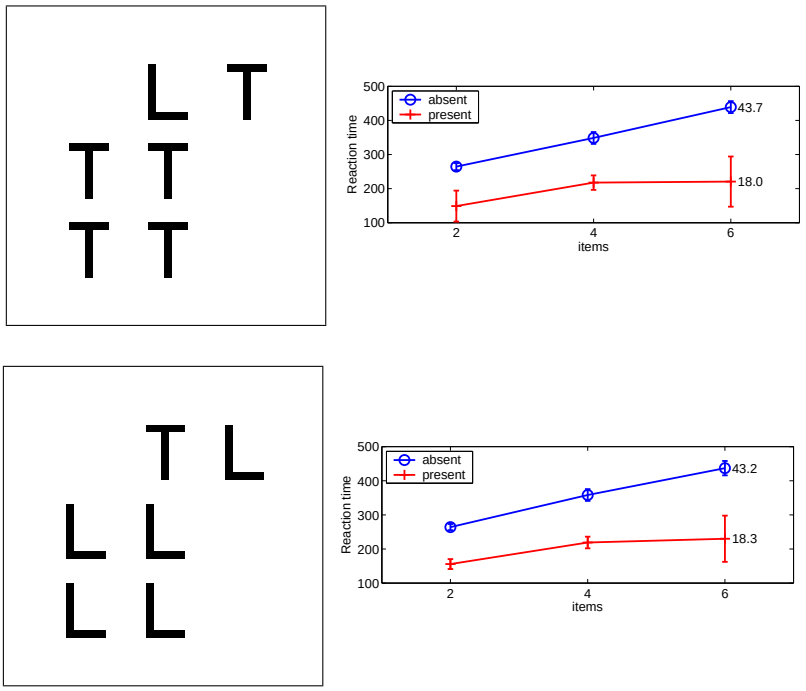

Figure 4. Two simulation results with "L" as target and "T" as distractor (top) and vice versa (bottom). The results are compatible with experimental evidence.

increases in the search functions and absent responses being lower than present responses. In these simulations each time a distractor was selected a recheck was performed with a given probability, in order to minimize target misses. Rechecking stoped entirely when either the target is found or all items were selected. The linear increase of SAIM's reaction time originates from two factors: First the time to select an item increased with the number of items, reflecting increased competition for selection. Second the number of rechecks increased with the number of items, since the probability of missing a target increases with the number of items. Fig. 5 shows the successful simulation of a search asymmetry [5], where a tilted line amongst horizontal lines is quicker than a vertical line amongst tilted lines. There is a bottom-up bias favouring the tilted line, which can only be overcome by top-down bias from the knowledge network. In SAIM the asymmetry stems from the fact that the time to select an item increases when top-down knowledge has to override a bottom-up bias. The increase results from the fact that the top-down bias from the knowledge network

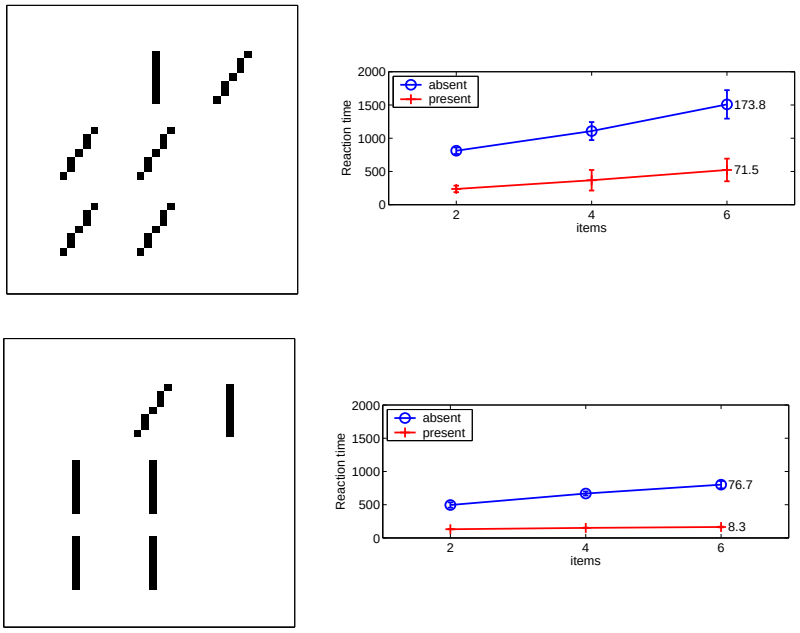

Figure 5. Simulation of a search asymmetry. Search for a tilted line amongst vertical lines is "parallel' (8.3ms/item), whereas search for a vertical line target amongst tilted lines produces a "serial" search (71.5 ms/item).

has to propagate through the contents network to the selection network. This leads to delayed resolution of activation in the selection network, after the network first follows its bottom-up preference to the distractor.

3.3. *Prediction and Experiment*

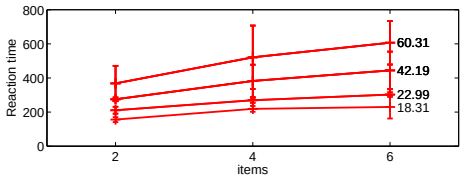

Figure 6. Effects of varying the initial activation of templates units for "T" as target and "L" as distractor. The results show that the search slope (numbers on the right to the graphs) decreased with increasing initial values.

In a series of simulations we tested the influence of the top-down bias

introduced by varying the initial activation values of template units in the knowledge network. Fig. 6 shows that the search slope decreased with increasing initial activation values. This pattern results from the fact that an increasing top-down bias leads to a decrease of the selection time and an increase of the hit-rate of finding the target, reducing the amount of rechecking. We assumed that the influence of the initial value can be construed as the influence of a cue preceding search displays; the cue "sets" initial values. Search is quicker when the cue "sets" the system for the target than when the cue "sets" an alternative item. We have confirmed this prediction in an experiment on human subjects [6].

4. Conclusion

We have demonstrated that SAIM can be successfully extended to include a feature extraction process. This new version can simulate typical results from visual search experiments. It also made a prediction about effects of top-down priming of search, that we confirmed empirically. The model may provide a powerful framework for understanding human search. In future work we will aim at replacing the present feature extraction by a more biological approach (e.g. using a Gabor filter) and at capturing recent evidence suggesting that grouping interacts with attentional processes [7].

References

1. D. Heinke and G. W. Humphreys. Attention, spatial representation and visual neglect: Simulating emergent attention and spatial memory in the Selective Attention for Identification Model (SAIM). *Psychological Review*, 110(1):29–87, 2003.
2. J. M. Wolfe. Visual Search. In H. Pashler, editor, *Attention*, pages 13–74. Psychology Press, 1998.
3. J. J. Hopfield and D.W. Tank. "Neural" Computation of Decisions in Optimazation Problems. *Biological Cybernetics*, 52:141–152, 1985.
4. E. Mjolsness and C. Garrett. Algebraic Transformations of Objective Functions. *Neural Networks*, 3:651–669, 1990.
5. A. Treisman. Features and Objects: The Fourteenth Bartlett Memorial Lecture. *The Quartely Journal of Experiment Psychology*, 40A(2):201–237, 1988.
6. D. Heinke, G. W. Humphreys, and C. L. Tweed. Testing the predicition of SAIM (Selective Attention for Identification Model). *Visual Cognition*, submitted.
7. J. Driver, G. Davis, C. Russell, M. Turatto, and E. Freeman. Segmentation, attention and phenomenal visual objects. *Cognition*, 80:61–95, 2001.