The Area Activation Model of Saccadic Selectivity in Visual Search

Marc Pomplun (marc@psych.utoronto.ca)
Eyal M. Reingold (reingold@psych.utoronto.ca)
Jiye Shen (jiye@psych.utoronto.ca)
Diane E. Williams (diane@psych.utoronto.ca)
University of Toronto, Department of Psychology
100 St. George Street, Toronto, Ontario, Canada M5S 3G3

Abstract

We present an approach towards a simple, explicit model of saccadic selectivity in visual search tasks. The model in its present state includes weights for target-distractor similarities and fixation field size as its only adjustable parameters. Based on these, the model predicts the statistical distribution of saccadic endpoints for any given visual search display. Besides providing an explicit and complete mathematical specification of the model, we demonstrate the performance of its computer simulation in a triple-conjunctive search task. The model successfully simulates empirical data reported by Williams and Reingold (in press).

Modeling Visual Search

How do we detect a prespecified target item among a set of distractors? Numerous studies employing the paradigm of *visual search* have attempted to answer this question (see Treisman, 1988 and Wolfe, 1998, for reviews). In a typical visual search task, subjects have to decide whether a search display contains a designated target item, indicating their decision by pressing either a "yes" or a "no" button. In most studies, reaction times (RTs) and error rates were analyzed as a function of the number of items in the display (display size). The majority of current models of visual search were based on data obtained within this paradigm.

An early attempt to model visual search is the Feature Integration Theory (Treisman & Gelade, 1980; Treisman, 1988). This theory proposes the existence of preattentive *feature maps*, one for each stimulus dimension such as color or shape. These maps are created in parallel after stimulus onset and allow immediate target detection if the target is defined by a unique feature in any single dimension (feature search). If the target is defined by a specific combination of features (conjunctive search), attention is necessary to locally combine the information of the corresponding feature maps. As a result, subjects have to inspect the display in an item-by-item fashion until target detection or exhaustive search. The Feature Integration Theory thus explains the finding that reaction time tends to increase with display size in conjunctive search tasks, while it is almost constant in feature search tasks.

A more recent approach is the Guided Search Model (Cave & Wolfe, 1990; Wolfe, Cave & Franzel, 1989; Wolfe, 1994), which proposes a two-stage model of visual search. In the first, parallel stage, an *activation map* containing likely target locations is created on the basis of both top-down and bottom-up sources of activation. The second, serial stage uses the activation map to guide visual attention from item to item, starting with the item with the highest activation, then proceeding to the second highest, and so on, until the target is found or the current activation falls below a certain threshold (see Chun & Wolfe, 1996).

Besides many variations of these two models, there are also more complex approaches like the one by Grossberg, Mingolla and Ross (1994). Their model uses artificial neural networks to achieve perceptual grouping of search displays into subregions. Visual search is assumed to proceed serially between these subregions and in parallel within them.

Recently, several researchers have analyzed participants' eye movements during visual search to supplement traditional RT and accuracy measures (e.g. Findlay, 1997; Hooge & Erkelens, 1999; Jacobs, 1987; Luria & Strauss, 1975; Motter & Belky, 1998; Rayner & Fisher, 1987; Scialfa & Joffe, 1998; Shen, Reingold, & Pomplun, in press; Viviani & Swensson, 1982; Williams, Reingold, Moscovitch, & Behrmann, 1997; Williams & Reingold, in press; Zelinsky, 1996; see Rayner, 1998, for a review). Some of these studies have further examined *saccadic selectivity*, i.e. the proportion of saccades directed to each distractor type, by assigning saccadic endpoints to the closest display item. Such studies have found a strong selectivity towards distractors sharing a particular feature with the target item (e.g. Findlay, 1997; Hooge & Erkelens, 1999; Luria & Strauss, 1975; Motter & Belky, 1998; Scialfa & Joffe, 1998; Shen, Reingold & Pomplun, in press; Williams & Reingold, in press; but see Zelinsky, 1996). Given that eye movements are usually accompanied by shifts of attention (see Hoffman, 1998, for a review), it seems that subjects can selectively attend to a critical subset of items in the display rather than perform an item-by-item search as suggested by the original Feature Integration Theory.

To date, no explicit model has been proposed which allows for simulating saccadic selectivity in visual search. In the present article, we propose such an approach, referred to as the Area Activation Model. Following the description of the model, we examine its performance by simulating the saccadic selectivity findings reported by Williams and Reingold (in press).

The Area Activation Model

The Area Activation Model is based on assumptions concerning three aspects of visual search performance: (1) the extent of available resources for processing, (2) the choice of fixation positions, and (3) the scan-path structure.

Processing resources -The extent of available resources for processing is determined by a two-dimensional Gaussian function with its peak centered at the current gaze position (e.g. Pomplun, Ritter & Velichkovsky, 1996). The standard deviation σ_f of the Gaussian function would be affected by a variety of factors such as task difficulty, item density, and item heterogeneity, but in essence should be a function of the area from which information is extracted during a fixation (henceforth "fixation field"). For example, if the target and distractors are easily discriminable and the density and heterogeneity of items are low, we would expect the fixation field to be larger than when discriminability is low and density and heterogeneity are high. This theoretical measure is likely to be correlated with the number or density of fixations in a given area. If the fixation field is smaller, we would expect more fixations per display area. In fact, in the current simulation we are using the empirically observed number of fixations per trial to adjust σ_f.

Fixation positions - Fixation positions are chosen to optimize the amount of information acquired. However, the execution of saccades entails a certain amount of error, which causes fixations to deviate from these optimal positions. Another source of error in empirical data is related to inaccurate measurement of eye movements. It is important for a valid comparison between empirical and simulated data to consider both saccadic error and measurement error.

For every point in the display it is possible to calculate its informativeness or relevance to the search task, creating an activation map. In the present simulation, we use weights corresponding to features along several dimensions to determine activation for individual items. A variety of models may suggest different activation maps (e.g. Cave & Wolfe, 1990; Wolfe, 1994).

In order to make the method transparent and applicable to a wide variety of tasks, we provide a general, explicit specification of the model. A search display consists of N items with positions (x_n, y_n) and features $f_n^{(d)}$ along D dimensions, $n \in \{1,\ldots, N\}$, $d \in \{1,\ldots, D\}$. The search target has the features $t^{(d)}$. Each dimension d is assigned a weight $w^{(d)}$, which currently has to be estimated on the basis of the results from a pilot-study. If, for example, subjects rely entirely on color, the color weight should be set to 1 and all other weights set to 0.

If an item n is identical to the target in dimension d, the item's feature activation $a_n^{(d)}$ is set to the weight of that dimension:

$$a_n^{(d)} = \begin{Bmatrix} w^{(d)}, \text{ if } f_n^{(d)} = t^{(d)} \\ 0 \ \ , \text{ otherwise} \end{Bmatrix}, \ n \in \{1,\ldots,N\},\ d \in \{1,\ldots,D\}$$

The total activation of item n is then calculated as the sum of its feature activations, implying the possibility of simultaneous guidance of attention by two or more dimensions:

$$a_n = \sum_{d=1}^{D} a_n^{(d)} \ , \ n \in \{1,\ldots,N\}$$

In a triple-conjunction search task, for instance, with color, shape, and orientation weighted 1, 0.5, and 0 respectively, a distractor item of the same color and shape as the target would receive a total activation of 1.5, surpassing those distractors with single-feature correspondence. Results from empirical studies support the hypothesis of combined activation across dimensions (see Williams & Reingold, in press).

As argued above, the activation map function $m(x, y)$ should reflect the amount of information that could be processed during a fixation at any position (x, y) in the display, given a Gaussian distribution of resources for processing. In the current model, $m(x, y)$ is calculated as the sum of total activations of all the items, with each item weighted by the amount of resources it receives, as a function of its distance from (x, y):

$$m(x, y) = \sum_{n=1}^{N} a_n \cdot \exp\left[-\frac{(x - x_n)^2 + (y - y_n)^2}{2\sigma_f^{\,2}}\right]$$

The fixation targets are chosen as those maxima (peaks) of $m(x, y)$ that are greater than or equal to the activation of a single target item, i.e. those coordinates (x_p, y_p) meeting the following two requirements:

$$\exists \varepsilon > 0 : \left|x - x_p\right| + \left|y - y_p\right| < \varepsilon \Rightarrow m(x_p, y_p) > m(x, y) \forall x, y$$

$$m(x_p, y_p) \geq \sum_{d=1}^{D} t^{(d)}$$

While the first requirement achieves a plausible selection of fixation points for most efficient acquisition of information, the second requirement simulates a subject's ability to give a negative response even before attending to every item in the display. According to this equation, subjects can decide whether a peak in the activation map is high enough to possibly contain a target item. They can thus stop the search after inspecting all relevant peaks, without directing their attention to the irrelevant ones.

After calculating the fixation targets, the actual fixation points are determined by simulating normally distributed saccadic error and measurement error. Saccadic error is assumed to increase with a larger fixation field, which corresponds to faster search, longer saccades, and a more diffused activation map. Accordingly, in the present simulation, we set the saccadic error parameter to equal the fixation field parameter σ_f. Measurement error is set to a constant standard deviation σ_m corresponding to the precision of the eye tracker used in the empirical study. The actual fixation point for an activation peak (x_p, y_p) is thus determined on the basis of the following probability distribution $p(x, y)$:

$$p(x, y) = \frac{1}{2\pi(\sigma_f^2 + \sigma_m^2)} \cdot \exp\left[-\frac{(x - x_p)^2 + (y - y_p)^2}{2(\sigma_f^2 + \sigma_m^2)} \right]$$

Scan paths - The structure of scan paths is governed by the principle that every fixation target, i.e. every relevant peak in the activation map, is visited exactly once. The order in which these fixation targets are inspected is chosen in terms of spatial optimization, as suggested by empirical results (e.g. Zelinsky, 1996). Among the unvisited peaks, the current implementation of the model always chooses the one that is nearest to the current gaze position. This type of local scan-path minimization - also termed the "Greedy Heuristic" - has been shown to to resemble human scanning strategies without assuming extensive planning processes (see Pomplun, Carbone, Koesling, Sichelschmidt & Ritter, submitted).

Turning back to the distinction between feature and conjunctive search, the current model makes the following predictions: If the distractors' activations are too weak to form peaks that exceed the target activation - for example, if the target has a unique feature in one dimension (feature search) - the target item produces the only relevant peak in the display, yielding a highly efficient "pop out" search. In contrast, increasing target-distractor similarity (e.g. conjunctive search) leads to more fixations and a stronger influence of display size on search performance. These predictions of the model are consistent with empirical results.

Empirical Validation of the Model

The Area Activation model is illustrated by simulating saccadic selectivity findings reported by Williams and Reingold (in press). The authors reported two visual search experiments with 32 participants in each experiment. Participants were presented with displays of 6, 12, and 24 items, half of them containing a target item defined by a unique combination of three dimensions - color, shape, and orientation. Each experiment consisted of a single-feature (SF) and a two-feature (TF) condition, in which the distractor items shared one or two dimensions respectively with the target item. While both experiments used the same colors (red and blue) and orientations (upright and rotated clockwise by 90 degrees), the stimuli differed in the discriminability of the shape dimension. Experiment 1 employed the similar letters E and F (low discriminability), whereas Experiment 2 used the distinct letters T and C (high discriminability). Figure 1 (upper row) presents a sample stimulus for each of the two experiments. Eye movements were measured with the SR Research Ltd. EyeLink system. The measurement error in this study was determined as σ_m =0.6 deg.

In our comparison of empirical and simulated data, only target-absent trials were analyzed in order to avoid the disruptive influence of target items (see Zelinsky, 1996). In the present article, only the results for display size 24 were simulated.

Since we had no a-priori knowledge about the subjects' fixation field in each of the four conditions (SF and TF conditions in Experiments 1 and 2), we used an iterative algorithm to adjust the model's fixation field parameter σ_f in such a way that the simulated number of fixations per trial matched the empirical one.

Another problem was to determine the weights $w^{(d)}$ for the color, shape, and orientation dimensions. We used the SF conditions in both experiments to adjust these weights and we tested their generality by applying them to the TF conditions. In the SF condition of Experiment 1, subjects showed strong saccadic selectivity towards color and equally low selectivity towards shape and orientation (see Figure 2, top row). This suggested that only the color dimension induced feature guidance, while shape and orientation were irrelevant to the search process. Consequently, for both the SF and TF conditions in Experiment 1, the weights were set to 1, 0, and 0 for color, shape, and orientation respectively. Experiment 2 differed from Experiment 1 only in the shape discriminability. Therefore, a larger shape weight was required in Experiment 2, but the other two weights had to be the same. We adjusted the shape weight to 0.6 in order to match the empirical saccadic selectivity towards the shape dimension in the SF condition of Experiment 2.

With these adjustments, the computer simulation of the Area Activation Model attempted to address several important questions: Is the model able to quantitatively reproduce the empirical saccadic selectivity? Does the implemented concept of simultaneous guidance by multiple dimensions match the human data, i.e. do the parameters for the SF conditions predict selectivity values in the TF conditions? Do the simulated gaze trajectories correspond to the empirical ones, as indicated by the distribution of saccade amplitudes?

Figure 1 (lower left) shows the activation map calculated by the computer simulation for the sample stimulus of Experiment 1. It reveals four peaks induced by groups of distractors sharing the target color blue, since in this condition only color features contribute to the activation map. As shown in Figure 1 (lower right), the simulation fixates once in the vicinity of each peak while always choosing the nearest unvisited peak as the next saccade target.

Figure 2 allows a comparison between simulated and empirical results, with each row referring to one of the four conditions. The first row shows a remarkable correspondence in the SF condition of Experiment 1, for both the amplitude and the feature selectivity of saccades.

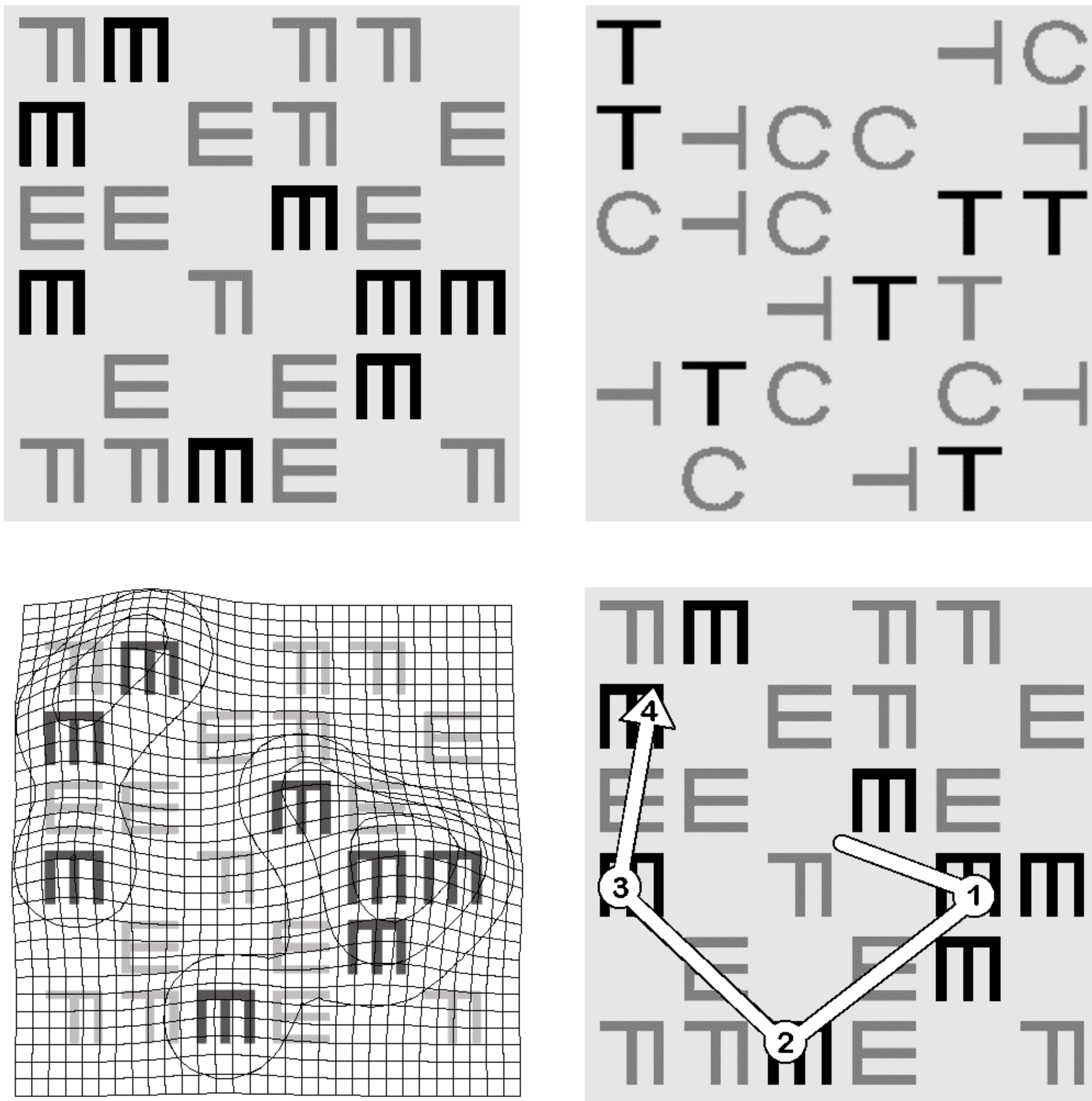

Figure 1: Sample stimuli and illustration of the Area Activation Model. Blue and red items are displayed in black and gray respectively. Upper left: Experiment 1, SF condition, target is a blue, upright “F” (absent). Upper right: Experiment 2, TF condition, target is a red, upright “T” (present). Lower left: Activation map ("activation landscape") calculated for the sample stimulus of Experiment 1. Lower right: Scan path generated by the model for the same stimulus. The four fixations correspond to the four peaks in the activation map.

The same is true for the TF condition, as shown in the second row. Despite a profound difference in search efficiency between these two conditions (3.77 versus 10.41 fixations per trial), the distribution of saccades and their selectivity is well predicted with the same set of parameters used in the SF condition.

With regard to the SF condition of Experiment 2, the model's saccadic selectivity once again closely resembles the empirical one, whereas the saccade histogram indicates a significant mismatch. The empirical data revealed a peak at an amplitude of approximately 3 degrees, but the model produced a smoother distribution extending further towards higher amplitudes. This discrepancy might be related to the high search efficiency in this condition (only 2.59 fixations per trial).

Finally, the TF condition, which is substantially less efficient (6.31 fixations per trial), showed an excellent correspondence between simulated and empirical data. The same parameters that failed to replicate the distribution of saccade amplitude in the SF condition almost perfectly reproduced the empirical amplitude histogram in the TF condition. Again, the model precisely predicted the effect of simultaneous guidance by two dimensions.

Experiment	1
Condition	SF

Fixations per Trial	3.77
Fixation Field Size	1.05 deg

Experiment	1
Condition	TF

Fixations per Trial	10.41
Fixation Field Size	0.96 deg

Experiment	2
Condition	SF

Fixations per Trial	2.59
Fixation Field Size	1.82 deg

Experiment	2
Condition	TF

Fixations per Trial	6.31
Fixation Field Size	1.06 deg

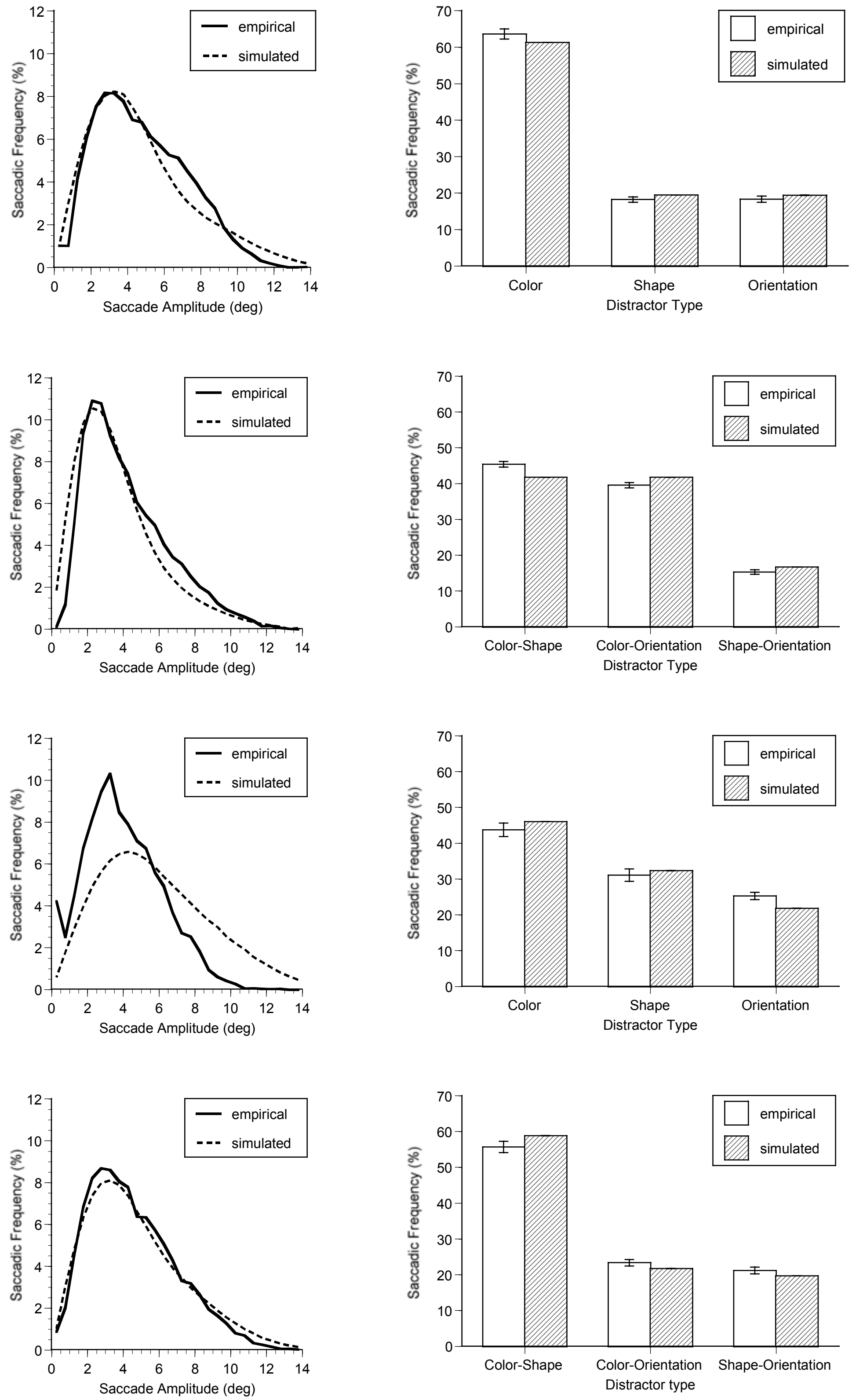

Figure 2: Comparison between empirical and simulated data with each row corresponding to one of the four experimental conditions. Left column: Empirical number of fixations per trial and simulated visual span size required to match the number of fixations. Middle column: Comparative histograms of saccade amplitude. Right column: Comparative diagrams of saccadic selectivity towards different distractor types.

Conclusions

In all four conditions, empirical saccadic selectivity was precisely replicated, supporting the concept of simultaneous guidance by multiple dimensions. Moreover, saccade amplitude produced by the model was remarkably accurate. One exception found was the SF condition in Experiment 2. This is perhaps due to the fact that search in this condition was highly efficient. It may be the case that highly efficient searches induce a qualitatively different saccadic scanning behavior. For example, if it is always possible to detect the target from the central gaze position, an efficient strategy could be to avoid any eye movements to the periphery. Another factor could be an increased amount of corrective saccades due to faster scanning of the display. Further research is necessary to investigate this issue.

As indicated by the model's accurate saccadic selectivity, not only the area-based activation map, but also the implementation of saccadic error - as identical to the fixation field size σ_f - have passed their first test. The generally successful replication of saccade amplitude supports the hypothesis of spatial scan-path optimization within the relevant display areas.

All in all, the current version of the Area Activation Model can be considered a promising approach towards an explicit, quantitative model of saccadic selectivity in visual search.

Acknowledgments

The preparation of this paper was supported by a grant to Marc Pomplun from the Deutsche Forschungsgemeinschaft (DFG) and a grant to Eyal Reingold from the Natural Science and Engineering Research Council of Canada (NSERC).

References

Cave, K.R., & Wolfe, J.M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225-271.

Chun, M.M., & Wolfe, J.M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology, 30,* 39-78.

Findlay, J.M. (1997). Saccade target selection during visual search. *Vision Research, 37*, 617-631.

Grossberg, S., Mingolla, E., & Ross, W.D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review, 101,* 470-489.

Hoffman, J.E. (1998). Visual attention and eye movements. In H. Pashler (Ed.), *Attention.* England UK: Hove.

Hooge, I. T., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Research, 39*, 1567-1575.

Jacobs, A.M. (1987). Toward a model of eye movement control in visual search. In J.K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition.* North-Holland: Elsevier Science Publishers.

Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception and Psychophysics, 17*, 303-308.

Motter, B. C., & Belky, E. J. (1998b). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805-1815.

Pomplun, M., Velichkovsky, B.M., & Ritter, H. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception, 25 (8)*, 931-948.

Pomplun, M., Carbone, E., Koesling, H., Sichelschmidt, L., & Ritter, H. (submitted). Modeling visual scanning strategies in two-dimensional object distributions. *Cognitive Science.*

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Rayner, K. & Fisher, D.L. (1987). Eye movements and the perceptual span during visual search. In J.K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From Physiology to Cognition.* North Holland: Elsevier.

Scialfa, C. T., & Joffe, K. (1998). Response times and eye movements in feature and conjunction search as a function of eccentricity. *Perception & Psychophysics*, 60, 1067-1082.

Shen, J., Reingold, E. M., & Pomplun, M. (in press). Distractor ratio influences patterns of eye movements during visual search. *Perception*.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology, 40A*, 201-237.

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

Viviani, P., & Swensson, R.G. (1982). Saccadic eye movements to peripherally discriminated visual targets. *Journal of Experimental Psychology: Human Perception and Performance, 8 (1),* 113-126.

Williams, D.E., Reingold, E.M. (in press). Attentive guidance of eye movements during triple conjunction search tasks. *Psychnomic Bulletin and Review*.

Williams, D.E., Reingold, E.M., Moscovitch, M., & Behrmann, M. (1997). Patterns of eye movements during parallel and serial visual search tasks. *Canadian Journal of Experimental Psychology*, *51*, 151-164.

Wolfe, J.M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*, 202-238.

Wolfe, J.M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-71). Hove, England UK.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419-433.

Zelinsky, G.J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research, 36*, 2177-2187.