Modeling the Control of Attention in Complex Visual Displays

Kelly S. Steelman-Allen
Jason S. McCarley
University of Illinois at Urbana-Champaign
2251 Beckman Institute
405 North Mathews Avenue
Urbana, Illinois 61801
ksteelm2@illinois.edu, mccarley@illinois.edu

Keywords:
Modeling, visual attention, display design

ABSTRACT: *A stochastic model of overt attention within a visual display or workspace is presented. The model integrates elements from several existing models of attention (Bundesen, 1987, 1990; Itti & Koch, 2000; Wolfe, 1994; Wickens et al., 2003) to provide predictions of the allocation of visual attention among discrete display channels and the number of eye movements needed to fixate the onset of a visual signal or event. The model was validated against data from an alert detection experiment (Nikolic, Orr and Sarter, 2004), with results demonstrating that the model can accurately predict the effects of color similarity, eccentricity, and dynamicity on attentional behavior and target detection.*

1. Introduction

In many operational domains, including aviation, nuclear power, and process control, one of the operator's primary tasks is to monitor for visual warnings or alerts. The detectability of such visual events is modulated by a variety of bottom-up and top-down factors, including the display context, the operator's mental model of system, and task demands. In a study by Nikolic, Orr, & Sarter (2004), for example, subjects monitored a display for the onset of a visual alert while engaged in a game of Tetris. Alert location and contrast, the presence of movement in the display, and the operator's level of attentional load were all varied. The detectability of alerts was found to depend on the interaction of these various factors, suggesting that design criteria that consider any one factor in isolation may not encourage effective display design.

The present paper describes a computational model to predict attentional behavior and target detectability within complex displays, offering designers a tool to test the effectiveness of various alerts in multiple display configurations and under varying task demands. The model incorporates elements from several computational models of basic attentional processes (Bundesen, 1987, 1990; Itti & Koch, 2000; Wolfe, 1994) within the heuristic SEEV framework of Wickens and colleagues (Wickens et al., 2003) to create a model of attentional behavior in dynamic environments**.**

2. The Model

The model assumes a scenario in which an operator monitors a display, comprising an array of discrete information channels, for some amount of time before the onset of a target event in one channel. The model predicts the steady-state distribution of attention among display channels, as measured in percentage of visual dwell time (McCarley & Kramer, 2006), prior to target onset; the likelihood of a scanning transition between any pair of channels prior to target onset; and the number of eye movements needed to fixate the target channel after the target appears. The model was implemented using Matlab 2008a and the Saliency Toolbox (Walther & Koch, 2006).

The model builds on the framework of Wickens' SEEV model (Wickens et al., 2003), which derives its name from the four forms of attentional influence that it posits: signal salience, the effort needed for attention to reach the signal, the operator's expectancy of the signal, and the task-relevance or value of the signal. The current model modifies and elaborates on the original SEEV model in multiple ways. First, it distinguishes between two forms of visual salience: *static salience*, based on local image-based feature contrast (cf. Itti & Koch, 2000), and *dynamic salience* (cf.Yantis & Jonides, 1990), based on moment-to-moment changes of static salience. Second, it distinguishes between two forms of top-down control: *channel prioritization*, based on the operator's estimates of the bandwidth and value of a given channel (cf., Senders, 1983), and *feature prioritization*, based on the operator's attentional set for a given color (cf. Wolfe, 1994). Third, it determines the salience of each channel computationally using the Itti and Koch (2000) salience model. Finally, it models the effects of effort on attentional scanning using a Gaussian spatial filter that simulates acuity loss in the peripheral retina and/or attentional tunneling, reducing the probability of long shifts of attention.

2.1 Inputs and model assumptions

As input, the model accepts image files of the pre- and post-target displays, a map of the display's information channels or *areas of interest (AOIs)*, and a parameter file specifying the bandwidth and value of each AOI. For simplicity, the model assumes that the pre-defined AOIs are the only locations in the image that can be fixated and that fixations always occur at the center of a given AOI. In its current form, the model also assumes that the target is noticed once it has been fixated, but this assumption could be easily replaced with the assumption of a probabilistic signal detection judgment.

2.2 Operation

The model operates by first producing a set of *base maps* representing various sources of attentional guidance. These maps are assigned *pertinence values* (Bundesen, 1990) based on the operator's task set, and the pertinence-weighted maps are averaged to produce a master map of attentional activation. Finally, a probabilistic choice model (Bundesen, 1987; Luce, 1959) determines the location of the operator's next fixation based on the attentional activation map.

2.3 Base Maps

The base maps represent four sources of attentional guidance: static salience, dynamic salience, channel priority, and feature priority.

Static Salience Map. The current model estimates the salience of each display channel using the computational model of Itti and Koch (2000). The Itti and Koch model employs center-surround filters to create a set of maps that represent feature contrast within the luminance, chromatic, and orientation dimensions. These within-feature contrast maps are then combined to form an overall saliency map, rendered in 16x16 logical pixels, with possible salience values ranging from 0 to a maximum value of 3. The current model normalizes the overall saliency map with respect to the maximum, to allow comparison of the static salience maps across simulations. For each iteration, *i*, of the model, a static salience map is generated based on the current display image. If the target has not yet appeared, then the pre-event input image is used to generate the map. If the change has already occurred, the post-event image is used.

Dynamic Salience Map. The dynamic salience map represents moment-to-moment changes in static salience resulting from the onset of the target or other sources of movement or flicker within the display. The model generates the map by calculating the Perceptual Euclidean Distance (PED) between the pre- and post-change images. The PED is similar to the traditional Euclidian distance but weighted to represent perceptual differences in color change detection for red, green and blue (Gijsenij, Gevers, & Lucassen, 2008). Calculating the PED for each pixel in the image produces a grey-scale map of changes in the display. This change map is then passed to the salience model, resulting in the dynamic salience map.

Feature Priority Map. The feature priority map is created by assessing the match between each pixel in the image and a set of target colors (e.g., red, green, blue, and amber). To assess the match for each color, the PED is calculated between the target RGB value and each pixel in the image. The color match is represented discretely, with a value of 1 indicating a match and zero otherwise. Pixels that fall within 40 units of the target color are considered a match. Each individual color map is then weighted according to its relevance to the task. For example, if red alerts represent danger and amber alerts represent potential danger, red may be assigned a value of 1 and amber a value of .75. The weighted color maps are then combined to form the final feature priority map.

Channel Priority Maps. The value and expectancy maps are both created heuristically. For each information channel in the display, the modeler provides the value and expectancy levels on a scale from 0-1. Both value and expectancy are assumed to remain constant during the task and are considered to be a function of the operator's mental model of the system and task. Accordingly, the values and expectancies are not considered model parameters that can be changed to better fit a set of data. Appropriate determination of the expectancy and values is thus an important step and requires the modeler to carefully consider both the nature of the display and the knowledge of the assumed operator.

2.4 Master Map

The master map of attentional activation values is created by averaging the activation of the base maps, with the input from each base map weighted by a pertinence value (Bundesen, 1990) assigned by the modeler. Pertinence values allow strategic changes in a modeled operator's attentional policy in response to changing task demands. For example, to allow attentional guidance driven entirely by bottom-up salience, the modeler can assign values of 1 to the static and dynamic salience maps and 0 to the other maps. Alternatively, to allow guidance based purely on top-down influences of bandwidth and information value, the modeler can assign a value of 1 to the two channel priority maps and 0 to the remaining three maps. Assigning equal pertinence values to all five base maps ensures that all five contribute equally to attentional guidance.

In order to simulate the effort required to execute a long attention shift (e.g., Ballard, Hayhoe, & Pelz, 1995) and/or the effects of acuity losses in the peripheral retina, a Gaussian spatial filter is applied to the master map at the center of the currently fixated AOI, L_i (cf., Parkhurst et al., 2002). The size of the filter, σ_{VL}, represents the size of the operator's visual lobe (Chan & Courtney, 1996) and can be adjusted to model individual differences (e.g., Pringle et al., 2001) or the influence of workload or stress (e.g., Atchley & Dressel, 2004) on attentional breadth.

2.5 Target selection

Finally, the mean activation level within each AOI is calculated to determine a single activation value, A_j, for each of the j AOIs. This value is the *attentional weight* of the AOI. The choice of an AOI for attentional selection is determined probabilistically based on the AOIs' relative attentional weights. More particularly, the probability that a given AOI is selected as the target for the next attention shift is given by a choice model (Bundesen, 1990):

$$P(\text{select } AOI_j) = A_j / \sum A,$$

where A_j is the attentional weight of AOI_j, and $\sum A$ is the summed value of the attentional weights for all AOIs. The choice equation effectively implements an independent race between AOIs for attentional selection (Bundesen, 1993)

To discourage consecutive attentional fixations on the same AOI, inhibition of return (IOR) can be applied to the attentional weight for the currently fixated AOI. IOR is a value between 0 and 1. In the case that IOR > 0, the attentional weight of the currently fixated AOI is multiplied by (1-IOR) before it is entered into the choice model, reducing the probability of a subsequent fixation in the same AOI. Thus, a value of IOR = 1 ensures that the model will never fixate the same AOI consecutively. Conversely, a value of IOR < 1 allows for consecutive fixations on a single AOI, introducing the possibility of attentional tunneling on channels of high bandwidth, value, or salience (Wickens & Alexander, 2009).

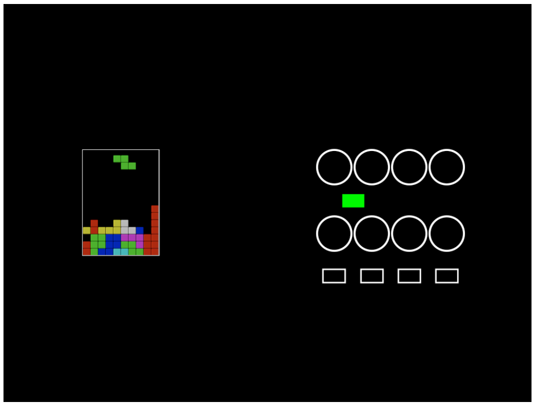
After the new fixation location is selected, a new master attentional activation map is created based on the current fixation location, and the selection process repeats. After the target event onset, the process continues until the model fixates the target AOI.

2.6 Model output

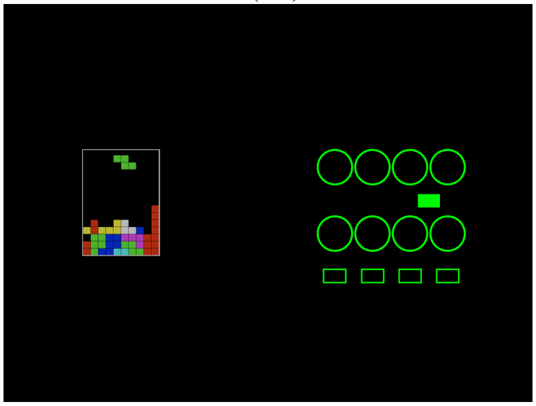
The model can be set to run for any number of fixations prior to the onset of the target, providing a distribution of steady-state scanning behavior within the pre-change display. After the onset of the target, the model continues to run until the changed AOI is fixated. Because the model is stochastic, the number of fixations required to locate the changed AOI varies between runs, producing a distribution of noticing times. This distribution can be used to predict mean cumulative target detection rate as function of time following target onset (Wickens et al., 2009).

3. Results

The model was validated against miss rates from an alert detection experiment (Nikolic et al., 2004). In the experiment, participants played a game of Tetris while simultaneously monitoring an adjacent display for the onset of a green alert. Three factors were manipulated in a 2x2x2 design: eccentricity of the alert with respect to the Tetris display (35 vs. 45 degrees of visual angle), color similarity between the alert and surrounding display objects, and dynamicity of objects near the alert. Schematic images from each of the eight conditions served as input to the model. Figure 3.1 presents the display for the low color similarity, near target location condition. Figure 3.2 illustrates the display image from the high color similarity, far target location condition. In the dynamic condition, the eight circular gauges contained random movement of the gauge pointer. In the static condition, there was no movement.



(3.1)



(3.2)

Figures 3.1 and 3.2 Representative displays from the low similarity, near target location condition (left) and the high similarity, far target location condition (right). Each display contained 15 areas of interest: 1 Tetris game, 8 gauges, 2 possible target locations, and 4 text boxes. The target was a green box, located between the two rows of gauges. In the low similarity condition, the objects surrounding the target were white. In the high similarity condition, the objects surrounding the target were green.

Pertinence values were assigned heuristically based on judgments about the relative usefulness of various forms of attention guiding information for detecting the target within each condition. More specifically, a pertinence value of 1 was assigned to each form of information that differentiated the target event from non-target events, and a value of 0 was assigned to all the remaining forms of information. Thus, for example, dynamic salience (due to the onset of the target) was assigned a pertinence of 1 in the static distractor conditions and 0 in the dynamic distractor conditions. Two experimenters independently assigned pertinence values for each condition and were in 100% agreement in all assignments (Table 3.1).

Source	High Similarity/ Dynamic	Low Similarity/ Dynamic	High Similarity/ Static	Low Similarity/ Static
Static Salience	0	1	0	1
Dynamic Salience	0	0	1	1
Value	1	1	1	1
Expectancy/Bandwidth	1	1	1	1
Attentional Set (Color)	0	1	0	1

Table 3.1 Pertinence values for each condition.

Note that the same sets of pertinence values were used in the near and far conditions. Distance effects were implemented by a Gaussian Spatial Filter with a standard deviation of 190 pixels, or approximately 15 degrees of visual angle. The IOR parameter was set to zero.

Pre- and post-alert images and the set of model parameters were input to the model. The model was run for 1000 iterations. Each iteration, the initial fixation was on a randomly selected AOI. After 100 fixations, the alert onset occurred, and the model was then allowed to run until the alert was fixated. To calculate a miss rate, the number of fixations-to-detection was first converted into a detection time by assuming a mean fixation duration. As the alert was assumed to remain visible for 10 seconds, if the detection time was greater than 10 seconds, that iteration was considered a miss. Accordingly, miss rates were dependent on the assumed fixation durations, with misses occurring after 10, 20, 30 or 40 fixations depending on whether 1000, 500, 333 or 250 ms fixations durations were assumed (corresponding to 1-4 fixations/second).

For each of the four assumed fixation durations, the Pearson correlation, Spearman's rank order correlation, and the root mean square error (RMSE) were calculated between predicted and actual miss rates (Table 3.2). Neither the Pearson correlation between the predicted and actual miss rates nor the rank order correlation varied significantly with the assumed number of fixations per second. The RMSE was minimized with assumed fixation durations of 250 or 333ms.

Fix/sec	r	r_s	RMSE
1	0.91	0.95	0.28
2	0.94	0.95	0.13
3	0.95	0.95	0.05
4	0.95	0.95	0.06

Table 3.2 Pearson correlation, Spearman's rank order correlation, and the root mean square error.

Figure 3.3 presents the predicted and observed miss rates for each condition, based on an assumed fixation duration of 250 ms. Figure 3.4 presents the same data collapsed across condition to illustrate the effects of target eccentricity, target color distinctiveness, and dynamic distractor content on predicted and observed miss rates. The model accurately predicted the empirical difference between the dynamic and static conditions, with moving gauges producing higher miss rates. The model also predicted the effects of both the eccentricity and color. As is evident in both figures, predicted miss rates generally underestimated observed miss rates (M_{diff}=-.042, SD=.043). Employing an assumed fixation duration of 333 ms helped to correct this effect, with underestimation of the miss rates in only 3 conditions, but overestimation in all others (M_{diff}=.025, SD=.432).

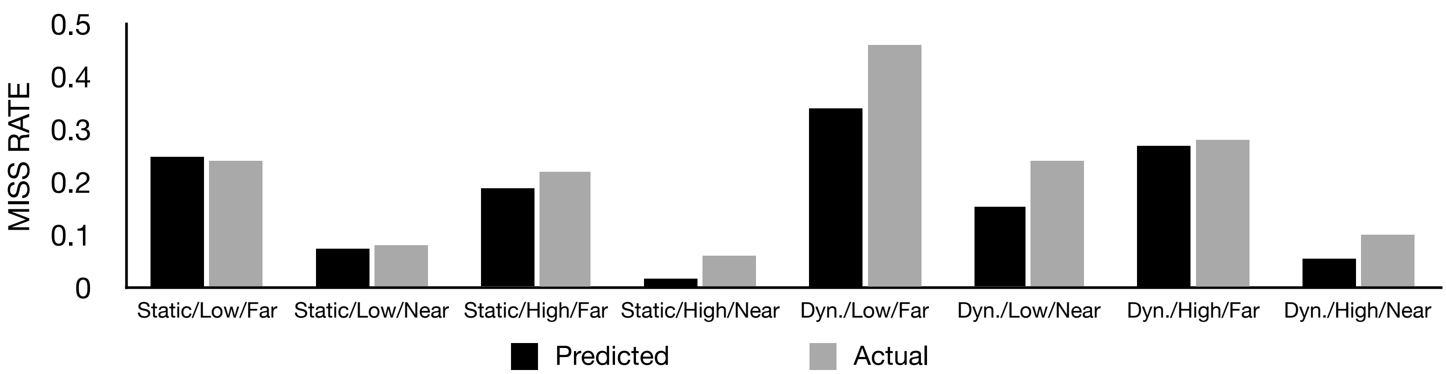


Figure 3.3 Predicted and actual miss rates for all 8 conditions.

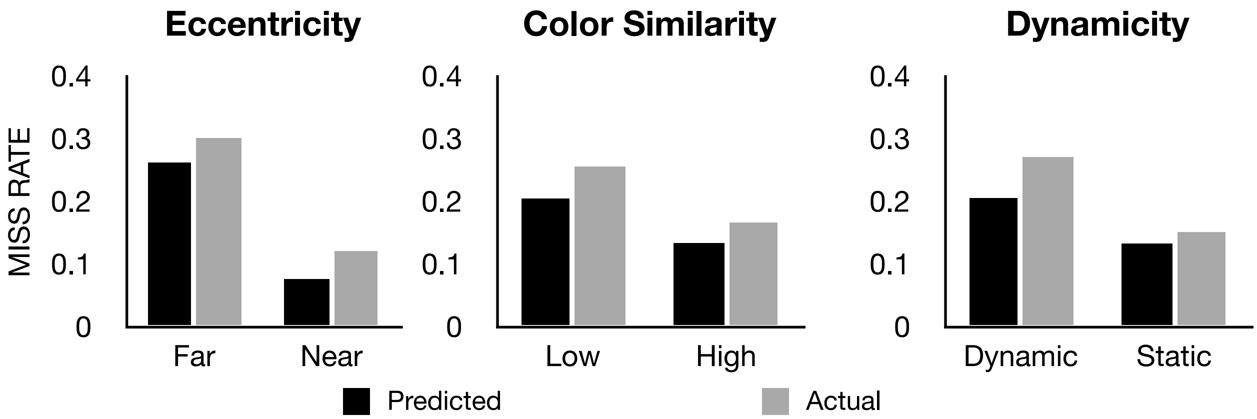


Figure 3.4 Predicted and actual miss rates, collapsed across conditions to illustrate the effects of eccentricity, color similarity, and dynamicity.

4. Conclusions

Based on the general framework of SEEV (Wickens et al., 2003), the current model assumes attentional guidance driven by signal salience, expectancy, and value, but distinguishes between static and dynamic visual salience and two manifestations of top-down guidance. The model thus accommodates multiple bottom-up and top-down factors that influence the noticeability of a visual event. It provides predictions of steady-state attentional behavior in a display and the number of eye movements required to fixate a visual event.

The model was validated here against miss rates from Nikolic et al.'s (2004) alert detection experiment. Results suggest that the model can reliably predict noticing behavior and can account for the effects of color similarity, eccentricity, and dynamic noise on target detection rates. Moreover, the validation confirmed that the model can be successfully fit using pertinence values selected through a simple heuristic. Additional validation is underway, focusing on modeling the distribution of oculomotor fixations within a complex workspace. Future efforts will attempt to model individual differences in attentional guidance and noticing, as well as the effects of mental workload on attentional behavior.

5. References

Atchley, P., & Dressel, J. (2004). Conversation Limits the Functional Field of View. *Human Factors, 46*(4), 664-673.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory Representations in natural tasks. *Cognitive Neuroscience, 7*, 66-80.

Bundesen, C. (1987). Visual attention: Race models for selection from multielement displays. *Psychological Research, 49*, 113-121.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*(4), 523-547.

Bundesen, C. (1993). The relationship between independent race models and Luce's choice axiom. *Journal of Mathematical Psychology, 37*(3), 446-471.

Chan, A. H. S., & Courtney, A. J. (1996). Foveal acuity, peripheral acuity, and search performance: A review. *International Journal of Industrial Ergonomics, 18*, 113-119.

Gijsenij, A., Gevers, T., & Lucassen, M.P. (2008) A perceptual comparison of distance measures for color constancy algorithms. Proceedings of the European Conference on Computer Vision, Volume I (pp. 208-221). Marseille, France: Springer.

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

McCarley, J.S. & Kramer, A.F. (2006). Eye movements as a window on perception and cognition. In M. Rizzo & R. Parasuraman (Eds.), Neuroergonomics: The Brain at Work. N.Y.: N.Y. Oxford University Press.

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why Pilots Miss the Green Box: How Display Context Undermines Attention Capture. *The International Journal of Aviation Psychology*, *14*(1), 39–52.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*, 107–123.

Pringle, H. L., Irwin, D. E., Kramer, A. F., & Atchley, P. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin and Review, 8*(1), 89-95.

Walther, D. & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks,19*, 1395-1407.

Wickens, C.D., & Alexander, A.L. (2009). Attentional tunneling and task management in synthetic vision displays, *International Journal of Aviation Psychology, 19, 1-17.*

Wickens, C., McCarley, J., Sebok, A., Steelman-Allen, K., Sarter, N. & Bzostek, J. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. Submitted to 2009 Annual meeting of the Human Factors & Ergonomics Society.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.

Wolfe, J.M. (1994). Guided search 2.0 A revised model of visual search. Psychonomic Bulletin & Review 1(2), 202-238.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 121–134.

Author Biographies

KELLY S. STEELMAN-ALLEN is a PhD student in Psychology at the University of Illinois at Urbana-Champaign. She has a M.S. in Human Factors from UIUC and Masters of Mechanical and Aerospace Engineering from Illinois Institute of Technology. Her research interests include basic and applied attention, display design, and instructional design.

JASON S. MCCARLEY is an assistant professor in the Department of Psychology and the Human Factors Division at the University of Illinois at Urbana-Champaign.