In Proceedings of the Sixth International Conference on Cognitive Modeling, 124-129.
Mahwah, NJ: Lawrence Earlbaum.

Explaining Eye Movements in the Visual Search of Varying Density Layouts

Tim Halverson (thalvers@cs.uoregon.edu)
Anthony J. Hornof (hornof@cs.uoregon.edu)
Department of Computer and Information Science, 1202 University of Oregon
Eugene, OR 97401-1202 USA

Abstract

This research advances computational cognitive modeling of visual search, and the synergistic relationship between cognitive modeling and eye tracking. The paper presents cognitive models of the perceptual, cognitive, and motor processing involved in the visual search of words in structured layouts that vary in density. The layouts are all-sparse, all-dense, or mixed. A principled approach is taken to account for eye movement data, specifically the mean number of fixations per trial and mean fixation durations. A random search strategy without replacement is used as a base model. The best-fitting model assumes that people examine two to three items per fixation regardless of the density. A new implementation of the EPIC cognitive architecture is used to build the models in this study. Modeling adjustments necessary to account for the data are discussed.

Introduction

Cognitive modeling is useful to the field of human-computer interaction because it reveals patterns of human performance at a level of detail not otherwise available to analysts and designers (as in Gray, John, & Atwood, 1993). The ultimate promise for cognitive modeling in human-computer interaction is that it provides the science base needed for predictive analysis tools and methodologies (Card, Moran, & Newell, 1983). This article reveals patterns of human performance in visual search, and contributes to predictive analysis of visual search.

The density of items in a display is one factor that has been shown to affect visual search. Bertera and Rayner (2000) varied the density of a fixed number of characters by varying the spacing between characters in a search task and found that search time decreased and the estimated number of letters processed per fixation increased as the density increased. Bertera and Rayner concluded that the effective field of view, the visual space from which information is perceived in a fixation, did not vary with the stimuli density. Ojanpää, Näsänen, and Kojo (2002) studied the effect of spacing on the visual search of word lists, and found that as the vertical spacing between words increased (i.e. as density decreased), search time also increased. In general, research examining the effect of density on visual search has found that more dense stimuli are searched faster per object, with a decrease in the number of fixations required to find the target being the largest factor influencing search time.

The modeling presented here focuses on the issues raised by previous research on density, e.g., the number of items perceived per fixation, and other fundamental perceptual and ocular motor issues of visual search. Previous modeling has used data from eye tracking to inform the development of models with respect to the order of search (e.g. Byrne, 2001; Hornof & Halverson, 2003). Here we use fixation duration and number of fixations to inform the development of other aspects of the models.

This paper presents models of a task that investigates the effect of local density on visual search. The purpose of these models is to determine the perceptual and ocular-motor constraints that are required to explain eye movement data collected with this local density task. Other aspects of the data, such as fixation order, are left for future research.

The Visual Search Experiment

The task modeled in this paper is the visual search of a known target among words in structured layouts. Figure 1 shows a sample layout from a trial. All layouts contained six groups of left-justified, vertically-listed black words on a white background. The groups were arranged in three columns and two rows. There were two types of groups of different densities: *Sparse* groups contained five words of 18 point Helvetica font. *Dense* groups contained 10 words of 9 point Helvetica font. Both group types subtended the same vertical visual angle.

There were three types of layouts: *sparse*, *dense*, and *mixed-density*. Sparse layouts contained six sparse groups. Dense layouts contained six dense groups. Mixed-density layouts contained three sparse groups and three dense groups. Figure 1 shows one such layout. The arrangement of the group densities in the mixed-density layouts was randomly determined for each trial.

Target and distractors items were words selected randomly from a list of 765 nouns generated from the MRC Psycholinguistic Database (Wilson, 1988). No word appeared more than once per trial. The words in the list were selected as follows: three to eight letters, two to four phonemes, above-average printed familiarity, and above-average imagability. Participants were precued with the target word before each layout appeared.

Each trial proceeded as follows: The participant studied the precue; clicked on the precue to make the precue disappear and the layout appear; found the target word; moved the cursor to the target word; and clicked on it.

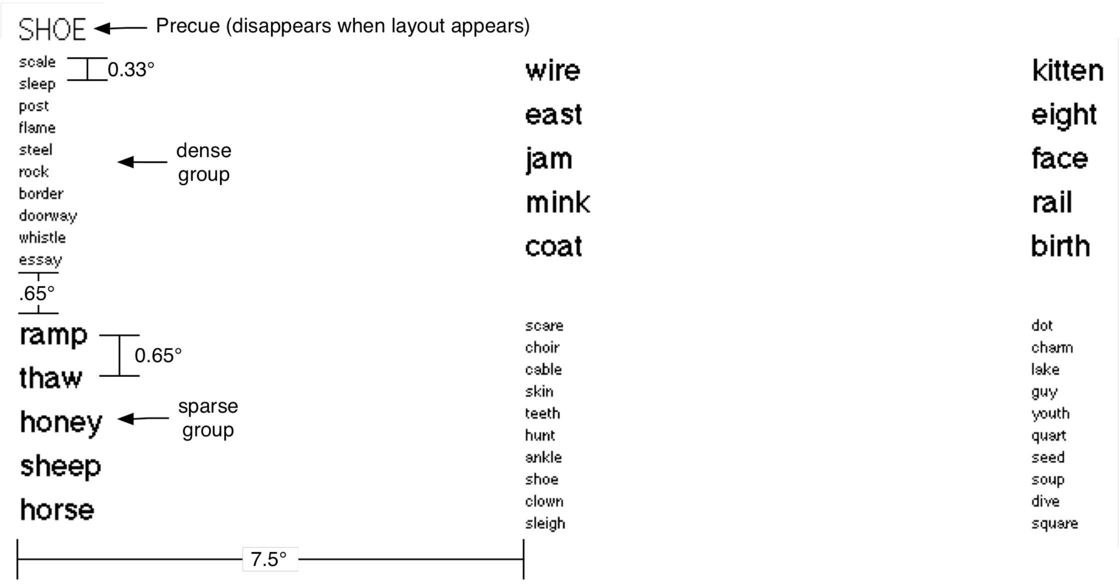

Figure 1. A mixed-density layout. All angle measurements are in degrees of visual angle.

The trials were blocked by layout type. Each block contained 30 trials, preceded by five practice trials. The blocks were counterbalanced using the Latin square technique. Eye movements were recorded using an LC Technologies Eyegaze System, a 60 Hz pupil-center/corneal-reflection eye tracker.

The Observed Data

The solid line plotted in Figure 2 shows the observed search time per trial, in Figure 3 the observed mean number of fixations per trial, and in Figure 4 the observed mean fixation duration. These data were collected from 24 participants (see Halverson & Hornof, 2004 for details). The mean number of fixations per trial and mean fixation duration are used for comparison with the model data for two reasons. First, they were both shown to vary significantly by layout type, F(2,46) = 60.17, p < .01 and F(2,46) = 61.82, p < .01, respectively. Second, the majority of the shift in search time per trial across the different layout types may be accounted for by the number of fixations and fixation duration. The search time per trial shows to what extent adjusting the number of fixations per trial and fixation duration can account for the search performance. The observed search time was also found to vary significantly with layout type, F(2,46) = 127.80, p < .01.

The data of primary interest are the eye movement data from the sparse and dense layouts. Previous work (Halverson & Hornof, 2004) suggested that the observed behavior for the mixed density layouts was affected by global strategies, which are not the primary concern in the present discussion. The important features in the observed data are that all three measures in this study, search time, number of fixations per trial, and fixation duration, increased when the local density increased.

It is well established that search time varies with distractor numerosity. However, the number of distractors cannot account for all of the shift in search time observed across the three layout types. If we normalize for the number of words per layout, the between-layout differences in mean search time per word and mean number of fixations per word are also significant, F(2,46) = 14, p < .01 and F(2,46) = 3, p = .05, respectively. However, search time per trial and the number of fixations per trial will be compared with the predicted data since they are more directly interpretable.

EPIC Cognitive Architecture

The models were constructed using the new C++ implementation of the EPIC (Executive Process Interactive Control) cognitive architecture (Kieras & Meyer, 1997). EPIC captures human perceptual, cognitive, and motor processing constraints in a computational framework that is used to build cognitive models. Into EPIC, we encoded (a) a reproduction of the task environment, (b) the visual-perceptual features associated with each of the screen objects, such as the text feature, and (c) the cognitive strategies that guide the visual search, encoded as production rules. These components were added based on task analysis, human performance capabilities, previous visual search model, and parsimony.

After these components are encoded into the architecture, EPIC executes the task, simulates the perceptual-motor processing and interactions, and generates search time and eye movement predictions. EPIC simulates ocular-motor processing, including the fast ballistic eye movements known as *saccades*, as well as the fixations during which the eyes are stationary and information is perceived.

The Models

Three models are presented in this section. Each improves on the last, correcting for a particular shortcoming in the prediction generated from the previous model. A good fit of the mean search time is the primary goal. However, while the mean search time is analyzed for each model, the models are improved incrementally based on eye movement data.

All models are based on a purely random, without-replacement, search strategy. While we do not necessarily assert that people move their eyes from item to item randomly, it may be that a random search strategy is a good first approximation for predicting *mean layout search time* without complicated strategies or visual features beyond the locations of objects. Such a strategy has the added benefit for *a priori* engineering models, as each object need be encoded with only one directly-extractable feature, its location. Subtler features, such as visual prominence or group-inclusion are not needed. Hornof (in press) found that a purely random search model with two to three items examined per fixation was good predictor for mean layout search time, even though it was a marginal predictor of search time per position.

Each model was run for 2,520 trials per layout type. The predictions generated by each model are discussed next.

Purely Random Base Model

The first model examined – the base model – was a purely random search model with all of EPIC's perceptual properties left at the default values. This base model is the standard to which we can compare subsequent models. In general, each word in the layout had an equal probability of becoming the destination of the next saccade. For this and following models, *text* is only available within one degree of visual angle (dov) from the center of fixation. We assume a without-replacement search. Any object for which the *text* had been perceived was excluded from being the destination point of future saccades.

The base model was overall a poor predictor of human performance. As seen in Figures 2 and 4, the predicted search times and fixation durations are incorrect in value and trend. It might seem that search time should decrease with dense layouts because more items fit into the fovea, but dense layouts also had more items to search.

Figure 3 shows the mean number of fixations. While the trend is incorrect for the predicted number of fixations as well, the prediction for the sparse layouts is quite good. This is promising and suggests that the purely random search model is a good starting point for modeling the characteristics of participant eye movements. As the predicted fixation durations have the greatest error of the two eye movement measurements, the next model will focus on improving the fixation durations produced by the model.

Wait-for-*text* Model

People appeared to adopt a search process that increased the duration of fixations on smaller, denser text. This could be achieved a number of ways in the model. One brute force approach would be for the production rules to directly set the fixation duration, though EPIC provides no such facility. Another would be to hold back each saccade until a certain amount of information is gathered from the currently fixated stimuli. This is straightforward to model in EPIC by (a) modulating the recoding time for a visual property (as in, dense text takes longer to recode) and (b) using a strategy that holds back each saccade based on this modulated

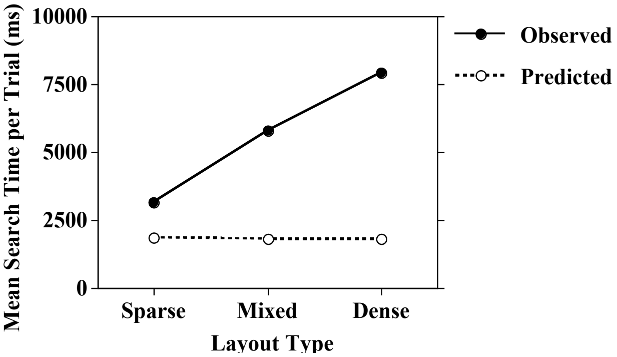

Figure 2. Mean search time per trial observed (solid line) and predicted (dashed line) by the purely random base model. Average absolute error (AAE) = 62.1%

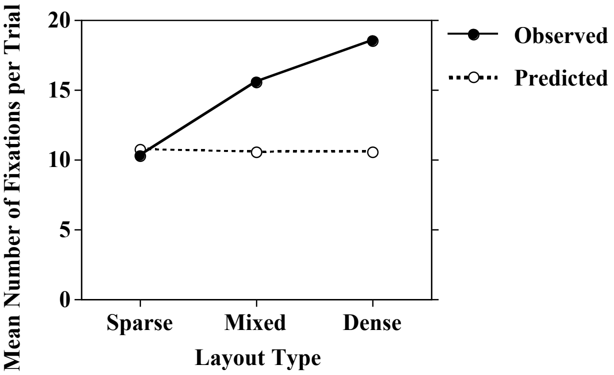

Figure 3. Mean number of fixations per trial observed (solid line) and predicted (dashed line) by the purely random base model. AAE = 26.7%

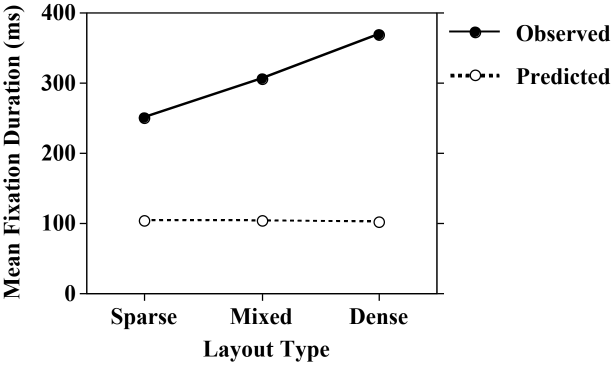

Figure 4. Mean fixation durations observed (solid line) and predicted (dashed line) by the purely random base model. AAE = 65.5%

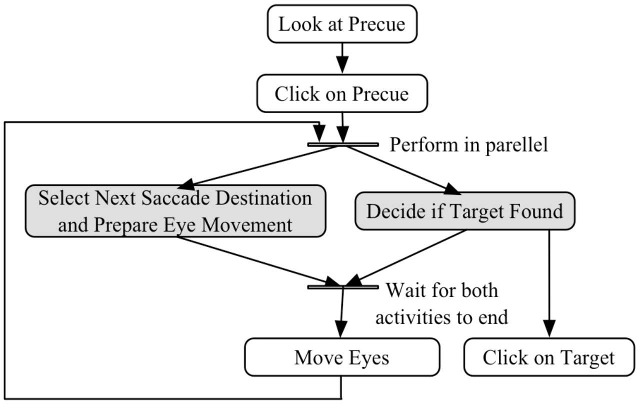

Figure 5. Flow chart of the Wait-for-TEXT strategy. Universal Modeling Language, a diagrammatic standard in software engineering, is used here.

feature. The Wait-for-*text* model uses such a strategy. Figure 5 shows a flow-chart based on the production rules.

The EPIC's default recoding time for the *text* feature is a constant 100 ms. This was modified when trying to explain the human data. As shown in Figure 4, the mean observed fixation duration was over 100 ms longer in the dense layouts than in the sparse layouts. To model this a stepped recoding function was introduced to calculate the perceptual time for a feature based on the proximity of adjacent items. If an object's closest neighbor was closer than 0.15 dov (a dense object), the *text* recoding time was 150 ms. Otherwise the *text* recoding time was 50 ms.

The base model initiated a saccade to the next randomly chosen object as soon as the previous saccade was complete. We hypothesized that humans may not initiate a saccade to another object before the text of the currently fixated object had been perceived, but that they may select another object and prepare the eyes to move to that object. Based on this belief, the procedure used for initiating saccades became: (1) Select an object at random as a saccade destination and prepare the eyes to move to that object. (2) Wait for the *text* of the currently fixated objects to become available before performing the prepared saccade. A similar prepare-then-perform strategy was used successfully in the modeling of another visual search task with EPIC (Kieras, 2003).

The predicted mean search time improved slightly with these modifications. As seen in Figure 6, there is now a very slight upward trend in the search time. However, the slope of the predicted search time line is not steep enough. The lack of a sharp increase between sparse and dense layouts correlates with the continued poor fit of the predicted number of fixations per trial. The model still does not make more fixations in layouts with dense objects. Further, the overall mean number of fixations has dropped in comparison to the base model. Though the mean fixation duration is now explained very well.

Detailed traces of the models revealed that the drop in the mean number of fixations was due to the prepare-then-perform strategy. The base model initiated approximately three additional fixations after the target had been fixated but before the *text* property for the target had become available. This resulted in roughly three more fixations per trial than the prepare-then-execute strategy, which inhibited additional fixations until the *text* property was perceived.

A large improvement was found in the predicted fixation durations. The inhibition of saccades and an increased recoding time for dense objects resulted in a differentiation in fixation durations similar to that in the observed data. The predicted data could have been further improved by reducing the *text* recoding time for sparse objects further, as

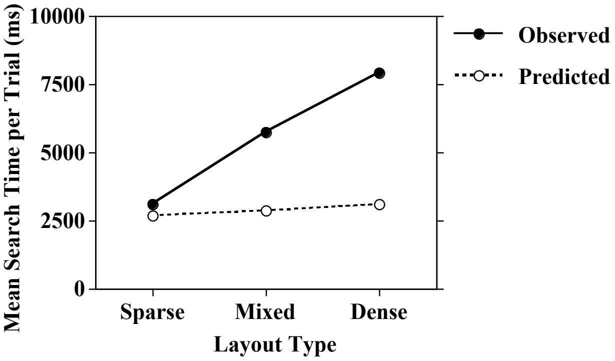

Figure 6. Mean search time per trial observed (solid line) and predicted (dashed line) by the Wait-for-the-TEXT model. AAE = 41.7%

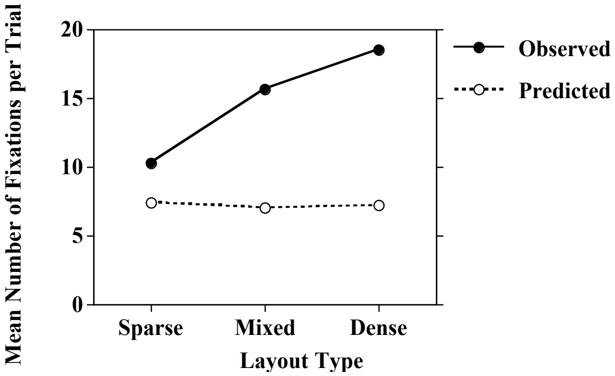

Figure 7. Mean number of fixations per trial observed (solid line) and predicted (dashed line) by the Wait-for-the-TEXT model. AAE = 48.1%

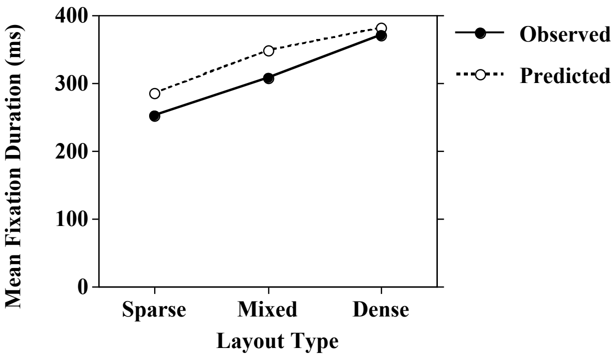

Figure 8. Mean fixation durations observed (solid line) and predicted (dashed line) by the Wait-for-the-TEXT model. AAE = 10.0%

the majority of the error in the predicted data lies in the sparse and mixed layouts. However, the purpose of this modeling was to approximate the ocular-motor behavior in the observed data, so further fine-tuning of fixation durations was not performed. Rather, since the greatest error now lies in the predicted number of fixations per trial the next model will focus on improving the number of fixations produced by the model.

Reduced *Text* Availability Model

A simplifying assumption in previous models directly influenced the predicted number of fixations. The assumption was that all text within the fovea (1 dov) is perceived for every fixation. This results in the model perceiving two to three sparse objects or five to seven dense objects in each fixation. Consequently, the model was able to perceive all *text* in a layout with an equal number of fixations, regardless of the layout density. The observed data suggests that humans do not do this. People require more fixations for dense text. An increase in the number of fixations required for dense objects can be achieved in a number of ways. One way is to reduce the region within which dense *text* can be perceived. Another is to reduce the probability of correctly perceiving *text* based on the size or spacing of the text. Both methods were tested in the models.

It was found previously that processing two to three items per fixation accounts well for the observed number of fixations in a search task (Hornof & Halverson, 2003). Here, we first limited the *text* availability to two to three items per fixations by decreasing the region of *text* availability for dense words to 0.5 dov. The default settings already limited sparse words to two or three per fixation. Perceiving two to three words per fixations resulted in a much better fit for the predicted number of fixations per trial. However, the model was still under-predicting the number of fixations per trial in all layouts.

A second availability function was created to vary the probability of perceiving the *text* property dynamically based on the distance to the closest neighboring object. This is one of several ways to measure density. One advantage of this measure is that it only requires the position of each item on the screen. If an object's closest neighbor was less than 0.15 dov away (a dense object), the probability of perceiving the *text* was 50%. Otherwise, the probability of perceiving the *text* was 90%. These probabilities were chosen because they would result in two to three items, on average, perceived per fixation across densities.

As seen in Figure 9, the predicted mean search time per trial improved considerably. Examining the ocular-motor behavior, we see that the predicted fixation durations improved slightly compared to the Wait-for-*text* Model, and the number of fixations per trial now closely approximates the observed data. The average absolute errors for the three measures range between 6.5% and 8.8%. This has been achieved by only modifying perceptual parameters and fundamental aspects of the strategy (i.e. prepare-then-perform).

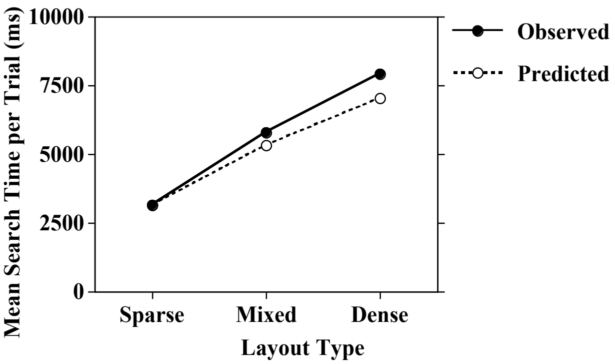

Figure 9. Mean search time per trial observed (solid line) and predicted (dashed line) by the Reduced Text Availability model. AAE = 6.5%

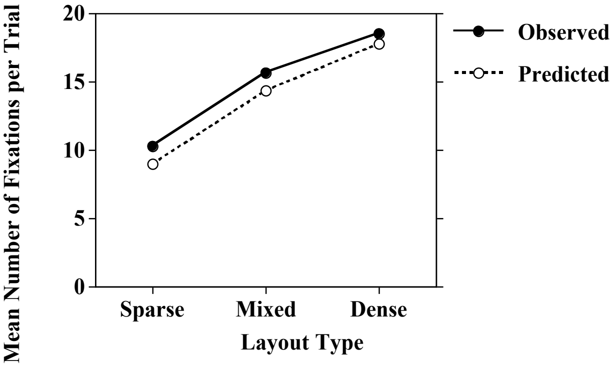

Figure 10. Mean number of fixations per trial observed (solid line) and predicted (dashed line) by the Reduced Text Availability model. AAE = 8.8%.

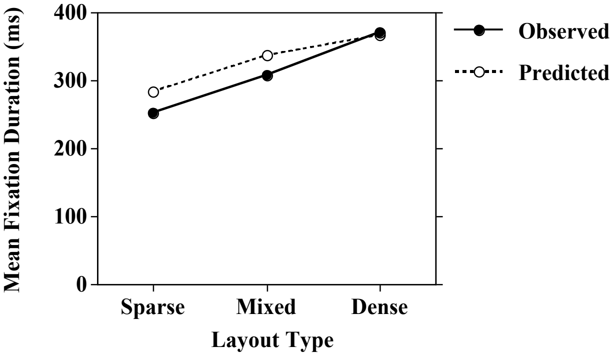

Figure 11. Mean fixation durations observed (solid line) and predicted (dashed line) by the Reduced Text Availability model. AAE = 7.8%

Discussion

Implications for cognitive modeling of visual search

A random search strategy is a reasonable first approximation that allows the analyst to focus on other fundamental ocular-motor activity that affects visual search. If an analyst can initially account for fundamental perceptual and ocular-motor activity with such a parsimonious model, the model may help to constrain the search space of more elaborate strategies.

One means of accounting for the number of fixations in a visual search of words is to limit the number of words perceived per fixation to two to three on average. Hornof (in press) found in that limiting the number of objects perceived per fixation to two to three items helped predict observed search times. The same assumption here helped to predict search time and number of fixations. However, it was not found that reducing the region of perception for denser objects was the best predictor, but rather maintaining the same region of perception and reducing the probability of perceiving denser objects so that an average of two to three items was perceived per fixation.

A straightforward and parsimonious means to account for fixation durations in a visual search of words is to use a prepare-then-perform strategy. The improvement in the predictions resulting from the use of this strategy suggests that future models that account for other aspects of the observed data should constrain the manner in which saccade destinations are selected. The saccade destination should be selected, and the eye prepared to move there, before the currently fixated text is perceived and the current fixation is complete. Otherwise, the predictions for saccade durations will probably be worse than was found here.

Implications for the theory of visual search

Bertera and Rayner (2000) concluded that the effective field of view did not decrease as density increased. The findings here work within that conclusion and expand upon it. It was found that if we tried decreasing the region in which *text* could be perceived (i.e. the effective field of view), that our model under-predicted the number of fixations required to find the target. However, if we left the region in which *text* could be perceived the same size, regardless of density, and changed the probability of perceiving *text* within that region, the models could better account for the observed data. The task modeled in this research differed from that used by Bertera and Rayner. In the current task, density was manipulated by varying the size of text and spacing (which is arguably more ecologically valid, see Halverson & Hornof, 2004). Still, similar conclusions were reached. Future work is required to study the effects of density where text size and spacing vary independently.

Conclusion

This paper presents models for a task that investigates the effect of local density on the visual search of words. A principled approach was used to account for the observed eye movement data. We started with a model that used the default perceptual parameters of EPIC and a random, without-replacement search strategy.

We found that a random search model that examines an average of two to three words per fixation predicts the observed number of fixations per trial data well. Further, it was found that a fixed region for *text* perception and a probability of perceiving *text* that varied with density was a better prediction than a variable region for *text* perception.

We also found that a model that waits for *text* to become available before initiating a saccade, but that prepares the next saccade in parallel with examining the *text*, is a good predictor of the observed fixation durations.

Further work is in progress to account for other observed data, such as the order of visitation and strategy shifts in mixed density layouts. The research presented here has established fundamental strategies and perceptual parameters that will constrain further modeling of the task.

Acknowledgments

This research is supported by the Office of Naval Research grant N00014-02-10440 and the National Science Foundation grant IIS-0308244. Both grants are to the University of Oregon with Anthony Hornof as the principal investigator.

References

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of effective stimulus in visual search. *Perception & Psychophysics*, 62(3), 576-585.

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, 55, 41-84.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction,* 8, 237-309.

Halverson, T., & Hornof, A. J. (2004). Local Density Guides Visual Search: Sparse Groups are First and Faster. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, New Orleans, LA.

Hornof, A. J. (in press). Cognitive Strategies for the Visual Search of Hierarchical Computer Displays. *Human-Computer Interaction.*

Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. *Proceedings of the Conference on Human Factors in Computing Systems*, Ft. Lauderdale, FL.

Kieras, D. E. (2003). Modeling Visual Search in the EPIC Architecture. *Presented at the meeting of Office of Naval Research Grantees in the Area of Cognitive Architectures*, University of Pittsburgh.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4), 391-438.

Ojanpää, H., Näsänen, R., & Kojo, I. (2002). Eye movements in the visual search of word lists. *Vision Research*, 42(12), 1499-1512.

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Usable Dictionary, Version 2. *Behavior Research Methods, Instruments, and Computers*, 20, 6-11.