Modeling the Visual Search of Displays: A Revised ACT-R/PM Model of Icon Search Based on Eye-Tracking and Experimental Data.

Michael D. Fleetwood and Michael D. Byrne
{fleet, byrne}@rice.edu
Rice University, Department of Psychology
6100 Main Street MS-25, Houston, TX 77005 USA

Abstract

As computer use becomes more visual in nature, researchers and designers of computer systems would like to gain some insight into the visual search strategies of computer users and the characteristics of displays that encourage the most efficient of these strategies. Icons, which are becoming increasingly prevalent, serve as the focus for a set of studies on the interaction of human vision with computer displays. Previous work (Fleetwood & Byrne, 2002) presented a study of "icon search" and a set of computational models of the task in the ACT-R/PM architecture. Presented here are an eye tracking study, conducted to examine the search strategies of users, and a revised model based on the results of the eye tracking study. The revised model incorporates EMMA (Salvucci, 2001) and changes in search strategy. Findings indicate key environmental influences of icon search (particularly set size and icon quality), evaluate the vision module in the underlying cognitive architecture, and provide some illumination on the strategies of users.

1. Introduction

Computer use has become a dominantly visual task. With the introduction of the graphical user interface (GUI), computer users spend less time trying to remember commands and operations and more time searching the screen for them. As such, much of a users' time is spent looking for and examining objects on their displays.

We would like to examine the interaction of human vision with GUIs in a common context, that of locating and selecting icons on a computer screen. In order to have value in an applied setting, we not only want to investigate the strategies of users in locating icons, but also the design characteristics of objects that will encourage the use of the most efficient of these strategies.

One of the primary issues we would like to examine is whether it is worth the time and effort to design "better" icons. Specifically, we would like to identify the features of icons that allow users to locate and select icons efficiently. Eventually, we would like to be able to make predictions regarding human performance in icon-based displays. To make such predictions, we need to develop an understanding of the search process that users go through when looking for a specific icon or group of icons.

The work described here builds on a previous set of experiments and computational models of human performance described in Fleetwood & Byrne, 2002. A brief discussion of the procedure and relevant results of that previous work is provided here. The work that builds upon those results, and which will be the focus of our discussion, is an eye tracking study of the task and a revised set of computational models informed by the eye tracking study.

2. Previous Research

We used ACT-R/PM (Byrne & Anderson, 1998) to model the experiment. Because the cognitive demands of the icon search task are minimal, modeling the perceptual-motor processes (e.g., shifting visual attention, pointing and clicking) with some fidelity is critical. The ACT-R/PM architecture combines ACT-R's theory of cognition (Anderson & Lebière, 1998) with modal theories of visual attention (Anderson, Matessa, & Lebière, 1997) and motor movement (Kieras & Meyer, 1997). ACT-R/PM explicitly specifies timing information for all three processes as well as parallelism between them.

In the basic structure of the experiment, participants were shown a target icon and corresponding filename and asked to locate and select the target icon amongst a set of distractor icons. The number of distractor icons was manipulated in set sizes of 6, 12, 18, and 24. The "quality" of the icons and distractors was also manipulated. Icons of "good" quality were based on the basic visual ("pop-out") features of shape and color, whereas lower quality icons had more complex shapes and were drawn in grayscale. Examples are shown in Figure 1. The dependent variable being measured was the response time of the participants. One potential independent variable that was held constant in this experiment was the number of icons matching the target in the search display. On each trial, one-third of the icons in the search display had the same pictorial icon and matching border. Thus, ultimately the user was forced to differentiate among the icons by the file name.

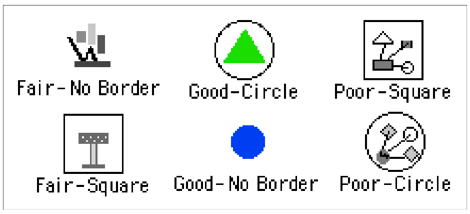

Figure 1. Examples of icons of good, fair, and poor quality used in the experiment. The good quality icons were each a single solid color, whereas the fair and poor quality icons were drawn in grayscale.

Examining that portion of the chart pertaining to the empirical data in Figure 2, it is evident that as icon quality decreases (good to fair to poor), response times increase. Also, not only are the three qualities significantly different, but the slopes of the lines representing each level of icon quality are reliably different. Also, revealed in the analysis was a reliable linear effect of set size. For each additional six icons added to the distractor set, the participants consistently took approximately 350 to 500 ms longer (depending on the icon quality) to locate the target icon.

The results of this experiment provide some insights into icon search. First, the cost in time of a search for a target icon is a linear function of the number of icons in the display. Second, an effect of icon quality was also produced, indicating that the quality of the icons has a significant effect on user response time. This serves to answer one of our original research questions, whether the effectiveness of an icon is significantly impacted by its relative level of quality.

2.1 Computational Modeling of the Experiment

Because our goal was to explore the space of strategies that users might employ, we constructed two models of the icon search task representing slightly different strategies.

Each of the models follow the same basic control structure. Where they differ is in the specific strategy that they use to search for and identify the target icon. In order for the model to select the target icon, it must first find an icon sharing some characteristic or "feature" of the target icon. (Each icon is represented in ACT-R/PM as a list of features. One of these features, e.g. gray circle, is randomly selected to guide the model in later search.) Then the model must read the filename below the icon and compare it to the target filename. One model, referred to as the "double-shift„ (DS) model is so named because it requires two shifts of attention to accomplish this process, one to an icon and one to the filename below it. The "text-look" (TL) model is so named because attention is focused directly on the filename below the icon, and the actual icon is never attended. As in the double-shift model, an icon sharing a feature with the target icon is located, but rather than shifting visual attention to the icon, it is shifted directly to the filename below the icon.

It is worth explicitly noting one prediction made by the double-shift model—the time that it predicts it will take participants to look at an additional icon. This is calculated as 420 ms—the time to run three productions (50 ms each) and two shifts of visual attention (135 ms each).

In the "good" quality icon condition, only icons exactly matching the target icon are examined by the model, and the RT (response time) by set size slope for the DS model should be 420 ms. This provides a basis for comparison with the experiment data.

2.2 Results

The fit of the model to the data for both the TL and DS models was encouraging (See Figures 2 & 3). Most importantly, both models have retained each of the pronounced effects that were seen in the data—those of set size and icon quality.

The proportion of variance explained (R^2) by the models relative to the data from the experiment is 0.99 and 0.98 for the DS and TL models respectively. The root mean square error (RMSE) and percent average absolute error between the models and data were 319.37 ms and 13.30% for DS model and 125.70 ms and 4.27% for the TL model.

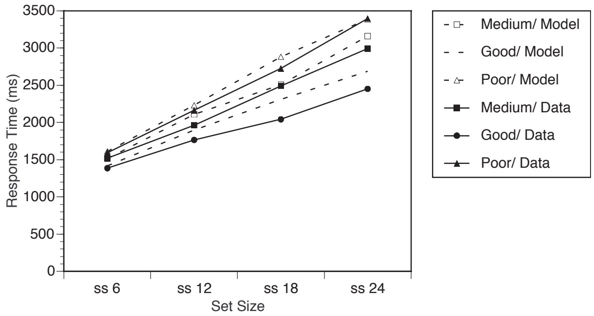

Figure 2. TL model comparison with data from the experiment.

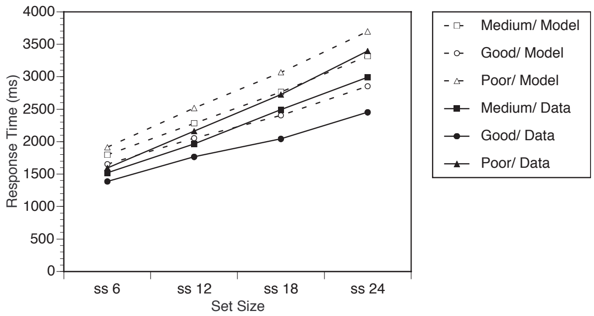

Figure 3. DS model comparison with data from the experiment.

Based on the standard fit metrics, the TL model appears to be a slightly better fit to the data than the DS model. This suggests that a search strategy which makes fewer shifts of visual attention may be more plausible than a strategy that makes two shifts of visual attention per icon examined.

Overall, we felt that the performance of the models relative to the experimental data was encouraging. However, one caveat deals with the predicted slope of the line for good quality icons across the four set sizes. As predicted, the slope for the DS model was calculated as approximately 420 ms. However, the slope for the participants in the experiment fell at approximately 355 ms. 420 ms does not fall within the 95% confidence interval for the 20 participants.

In this instance we have found an aspect of the model that does not conform well to the data. Real subjects can find the icon faster than our model. In order to examine how the search strategies of participants differed from the model strategies, we studied users' eye movements while engaged in the task, accomplished through the use of an eye tracker.

3. Eye Tracking the Icon Search Task

Eye tracking has been used by researchers to make fine distinctions regarding the processes used in a visual search task (for example, Zelinsky & Sheinberg, 1997; Findlay, 1982) and to gather information on the features of objects that drive visual search (Williams, 1967; Gould & Dill, 1969). In the realm of human-computer interaction, it has been used as a means of understanding the processes underlying the behavior of computer users (e.g. Byrne, 2001; Ehret, 2002).

This eye tracking study was designed to provide some insight into the processes underlying icon search and the features of the display driving that search.

3.1 Methods

3.1.1 Participants

The participants in the experiment were 10 undergraduate students at Rice University.

3.1.2 Design

The design of the experiment was nearly identical to the previously described experiment. The independent variables manipulated were set size (i.e. number of icons in the distractor set; 6, 12, 18, or 24), target border (the target icon could have a circle, square, or no-border), and icon quality (good, fair, poor). The dependent variable measured was reaction time, and with the additional use of an eye tracker, the participants' eye movements were recorded.

3.1.3 Apparatus/Materials

The eye tracker used was an ISCAN RK726/RK520 HighRes Pupil/CR tracker with a Polhemus FASTRACK head tracker. Head-mounted optics and a sampling rate of 60 Hz were used in the experiment. Point of regard (POR, also referred to as point of gaze) and cursor position were recorded. Where and when fixations occurred was calculated using a dwell-based technique.

3.1.4 Procedure

Participants were presented with a target icon and its corresponding file name. Clicking a "Ready" button presented them with a screen that contained a number of icons (6, 12, 18, or 24), one of which was the target icon. The participant's task was to identify the target icon and click on it as quickly as possible. Clicking on an icon brought them to the presentation of a new target icon. Response time was measured from the time they clicked on the ready button to the time they clicked on an icon in the distractor set.

The location of the target icon was randomly selected for each trial. The file names were randomly selected without replacement from a list of 750 names. Once exhausted, the list was recycled.

Each participant completed four blocks of 36 trials in addition to the practice block for a total of 180 trials. Each independent variable was examined at each level of the other independent variables (4 x 3 x 3 = 36). The order of presentation was randomized.

3.2 Results

Patterns in the fixation data were similar to those found in the reaction time data from the previous experiment—i.e. as set size increases and icon quality decreases, the average number of fixations increases (as does response time). There were reliable main effects of set size, $F(3, 27) = 77.08$, $p < 0.001$, and icon quality, $F(2, 18) = 56.60$, $p < 0.001$.

In Figure 4, the proportion of target fixations to total fixations is presented as a function of icon quality and set size. Target-matching icons were those icons identical to the target icons, which comprised one-third of the distractor set. Participants had a higher proportion of target fixations relative to non-target fixations with better quality icons, $F(2, 18) = 7.87$, $p < 0.01$, with Huynh-Feldt correction. Participants made a higher proportion of fixations to target-matching icons than would be expected if fixations were randomly directed, $t(9) = 6.90$, $p < 0.01$.

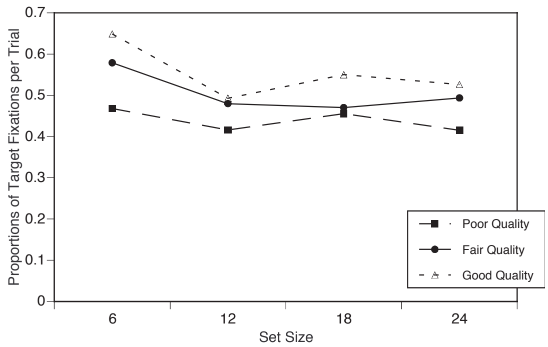

Figure 4. Proportion of target-matching fixations to total fixations by icon quality and set size, indicating that participants made a higher proportion of target-matching fixations with better quality icons.

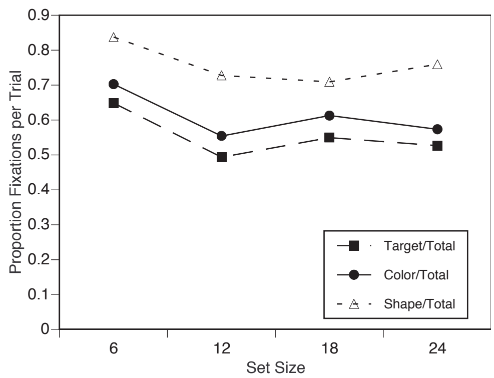

Figure 5. Proportion of specific fixations to total fixations.

Because the models make several predictions regarding the reaction time of participants in the good icon quality condition, data pertaining to this condition are examined separately and in greater depth (See Figure 5). Target-matching fixations and two other "types" of fixations were examined. Color fixations are those fixations on an icon of the same color as the target icon, regardless of shape. Shape fixations are fixations landing on an icon of the same shape as the target icon, regardless of color.

We were interested in whether the proportion of non-target fixations that were directed at an icon of the same color or shape as the target was reliably different than the proportion that could be expected if the fixations were random (0.09 and 0.45 for color and shape respectively). For color fixations there is an indication that the search strategies of participants were at least in part driven by the color of the icon, $t(9) = 2.77$, $p < 0.05$. For shape fixations, there is no such indication, $t(9) = 0.25$, $p = 0.81$.

Several qualitative patterns emerged in the data, which are difficult to quantify, but are nonetheless informative. For one, it seemed that participants in the experiment used different search strategies depending on the level of quality of the icons. For instance, in the good quality condition (Figure 6), the search strategy used by participants was often directed specifically at the icons matching the target icon. In this case, the saccades were nearly all directed to a target icon or fell in the area between two groups of target icons, leaving whole areas of the distractor set unexamined. Further, this "directed" strategy often worked with a "grouping" strategy where participants began their search with the largest group of matching icons and proceeded to smaller groups of matching icons until the target was identified. In contrast, search strategies in the poor quality condition were not directed at icons matching the target icons and might cover the whole set of icons, possibly in a circular or zigzag pattern.

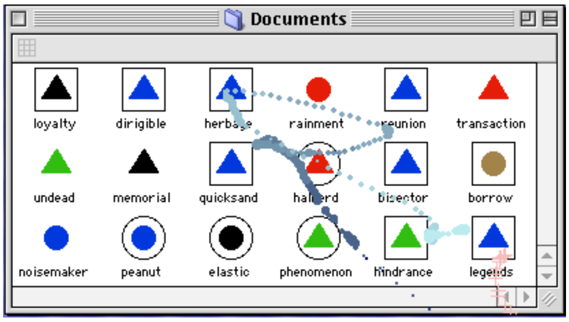

Figure 6. Example of a directed search with good quality icons. The round dots indicate point of regard, going from darker to lighter with time. The cross-hairs (in the lower right) indicate the position of the mouse. (The target, a blue triangle, is associated with the icons labeled dirigible, heritage, quicksand, reunion, bisector, and legends, which is the target) The participant begins with the largest group of matching icons and eventually proceeds to the single target-matching icon in the lower right.

3.3 Discussion of Eye Tracking Results

The data revealed that participants were more accurate in their search with better quality icons. This effect was manifested in the proportion of target fixations to total fixations, which increased with each level of improvement in icon quality. There was also some evidence for this effect at a qualitative level, manifested in the "directed" search strategies in the good quality icons and the "undirected" search strategies seen with poor quality icons.

The eye tracking data also provided us with some information as to which features of the good quality icons were used by participants to guide their "directed" search. Participants made a higher proportion of fixations to non-target icons of the same color as the target icon than would be expected if non-target fixations were randomly directed, indicating that color is a feature guiding search in the good quality condition.

4. Revising the Model

It was clear from looking at the search patterns of participants (Figures 6) that the TL strategy was a closer approximation of their search strategies. However, even the Text-Look model falls short of the strategies of the experiment participants, who often do not even look directly at a filename to determine whether it is the target filename.

Another area for improvement is in the number of shifts of visual attention made by the models. Both the DS ant TL models made many more shifts of visual attention (7.55 and 5.60 average shifts per trial, respectively) relative to the average number of fixations of participants per trial (3.23). Again, the Text-Look model is a closer approximation of the participants' performance; however even the TL model makes substantially more overt shifts of visual attention than participants.

This leads us to consider an issue in the underlying cognitive architecture of ACT-R/PM that other authors have discussed previously (Salvucci, 2000). ACT-R/PM currently assumes a direct correspondence between unobservable attention shifts and observable eye movements; that is, people fixate the target of attention. Such an assumption holds in some cases, but it is agreed upon in the research community that it does not hold in general (Henderson, 1992; Rayner, 1995). In order to address this issue, we turned to a computational model of eye movements that improves upon some of the underlying assumptions of ACT-R/PM's vision module.

4.1 Eye Movements and Movements of Attention (EMMA)

EMMA (Salvucci, 2000) serves as a bridge between observable eye movements and the unobservable cognitive processes and shifts of attention that produce them. Concerning eye movements, the model describes whether or not eye movements occur, when they occur, and where they land with respect to their targets. Concerning visual encoding, the model describes how peripheral viewing and object frequency affect the time needed to encode a visual object into an internal representation. Essentially, encoding time increases as object eccentricity increases and as object frequency decreases.

The incorporation of EMMA was expected to have two primary effects on the models. First, it was expected to reduce the number of shifts of visual attention. When the encoding time for a visual object is less than the time to make the labile portion of the eye movement to that object, the eye movement is not made, even though the object has been examined. Second, by calculating the time to make an eye movement, EMMA allows us to make predictions as to the relative efficiency of various icon search strategies. Specifically, strategies that make shorter shifts of visual attention can be expected to be more efficient.

4.2 Improving the Model Search Strategies

Our goal was to change the search strategies used by the model based on the results of the eye tracking study. For example, there was some qualitative evidence for a "grouping strategy" in the eye tracking study, particularly with good quality icons. As a result, we chose to implement a simple strategy to account for this behavior in our new models. The model would simply select the target-matching icon nearest to the icon that is the current focus of visual attention. Thus, the models would search within a group of target-matching icons before moving on to a new group. Such a strategy also ties in with the predictions made by EMMA. Specifically, a strategy that makes the shortest possible shift will be the most efficient strategy.

5. Modeling Results

The performance of the revised model was comparable to that of the previous versions. For the TL model; the RMSE was 129 ms; the PAAE was 5.89%, and the R^2 was 0.99 (Figure 7). The slope of good quality line was approximately 300 ms.

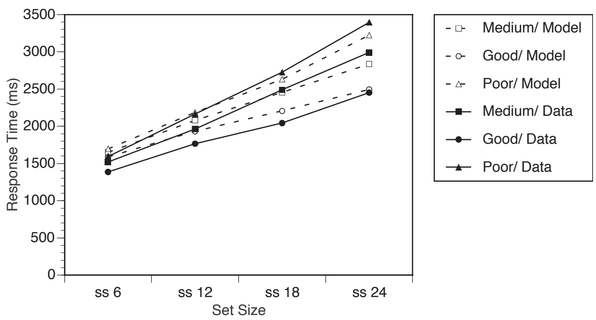

Figure 7. Response time by set size and icon quality for the revised Text-Look model and the experiment data.

We also found that the qualitative performance of the model was quite improved with respect to the "directed" and "grouping" strategies seen in the data. An example of a trial where these strategies were employed was given, which was shown in Figure 6. As an example of the capability of the models, the exact same trial was run with the model (see Figure 8). The line running through the figure shows the resulting trace of the POR data of the newest TL model. The model begins its search from the "Ready" button and enters the depicted portion of the trial from the lower-right corner. The model proceeds in a fashion quite similar to that of the human participant, first examining the largest group of target-matching icons before moving on to the nearest group and finally to the target icon in the lower-right corner of the window (labeled "legends") The model would not follow this exact search pattern every time (the first icon that the model selects for search is randomly chosen from the target-matching icons presented), but the capability of the model to mimic human performance in this respect is encouraging.

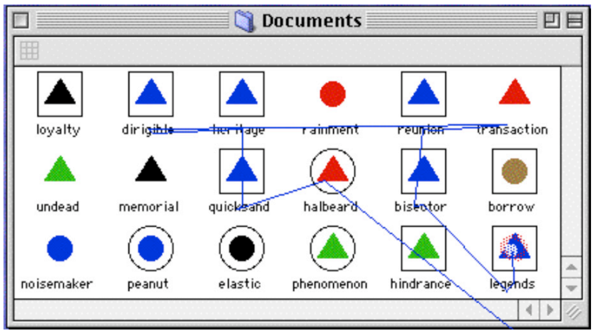

Figure 8. Example of the Text-Look model running an identical trial to that presented in Figure 6. The line indicates the POR path of the model. The model POR data begins at the "Ready,, button, which enters the view in the lower-right corner and finishes by selecting the icon above the filename "legends. "

5.1 Discussion of Modeling Revisions

The effect of only incorporating EMMA into the models was an overall increase in response time. The previous models used a constant parameter of 135 ms for each shift of visual attention. EMMA uses a set of algorithms to compute the encoding and saccade time A close examination of this attribute of EMMA revealed that longer saccades, such as those from one side of the distractor set to the other side, took an estimated time much greater than 135 ms, and were thus responsible for much of the increase in average saccade time.

Because longer saccades result in much longer response times, one way to reduce the overall response time of the model is to reduce the average length of saccades that it must make. The "nearest" strategy that was implemented was a nearly optimal strategy for minimizing saccade length.

This "nearest" strategy has further implications as well. For one, it adds some credence to our observation that participants seemed to search by groups of target-matching icons, which we were only able to verify at a qualitative level. Such a strategy would generally result in a very short average saccade distance.

The nearest strategy also has implications well beyond the realm of icon search. Tullis (1997) states, "The ways in which the elements are grouped plays an important role in both the ease with which the users can extract the information and the interpretations that they assign to it." Other researchers have made similar distinctions (Cakir, Hart, and Stewart, 1980). From this perspective, the organization of information on the screen has value to the user by giving them some additional categorical information as well as improving the general "readability" of the information. Not to diminish the value of this categorical information, but from the perspective of our modeling effort, grouping the information on the screen adds value at a much lower level in the cognitive system—in the visual search strategies employed by users. Grouping information will reduce the number and average distance of saccades made by the user while searching for a desired piece of information. Shorter saccades and less of them will result in finding the desired information more quickly.

It is likely we could improve the model to experiment data fit beyond its current state if we manipulated the feature lists that comprise how ACT-R/PM "sees" each icon. Each icon is represented in ACT-R/PM by a list of features, which were constructed for the initial models and subsequently tweaked to achieve a good model to data fit. It is encouraging that the revised models showed similar performance to previous models without any changes to the feature lists. However, the fact that construction of the feature sets plays such an important role in the predictive power of the model indicates a clear weakness in modeling a visual process in ACT-R/PM. In fact, this issue goes beyond ACT-R/PM; to our knowledge, no one has developed a method for systematic feature decomposition of displays such as those used here. Much of the predictive power of modeling in general is lost when some of it has to be done in a post hoc manner. Without this work, much of modeling the visual world will functionally remain a "free parameter" in any modeling effort.

6. General Discussion and Conclusions

Overall the performance of the models was encouraging. Each of the major trends in the data was well captured, the effect of set size, the effect of icon quality, and the general icon search strategies employed by users.

The number of icons in the display is a powerful predictor of icon search performance. In each of the experiments and in each of the models, there was a linear increase in icon search time with an increase in set size.

The level of icon quality also proved to be a powerful predictor of performance. Designing effective icons adds value to a system by reducing user search times.

The strategies that were implemented in the models also provided some insight into the strategies of human computer users. We developed two different strategies of icon search and were able to achieve comparable results with both strategies. We also found that the features of the display, such as set size and icon quality, affected both models in similar fashions. This suggests that icon search is a cognitive process that is driven in a "bottom-up" fashion more than a "top-down" fashion—i.e. variation in the characteristics of the display, such as the quality and number of the icons, have a greater impact on the search times of users than does variation in the strategies of users.

Our original goal in this set of studies was to gain some understanding of the icon search process. This set of studies sheds some illumination on the factors that contribute to that process. A larger goal of this line of research is to develop an engineering level model of the task that can aid in the design of real world systems. The models we have developed here make a lot of headway in that direction and have served to point out where future cognitive models and modeling architectures (ACT-R/PM in this case) need to be improved. More importantly, the models developed here serve to make some initial predictions regarding the performance of users in icon based systems, information that we hope will aid the designers of visual displays.

References

Anderson, J. R., and Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation of visual attention. *Human-Computer Interaction*, 12, 439-462.

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

Byrne, M.D. & Anderson, J.R. (1998). Perception and action. In J.R. Anderson & C. Lebiere (Eds.), *The Atomic Components of Thought*, 167-200. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cakir, A., Hart, D. J., & Stewart, T. F. M. (1980). *Visual display terminals: A manual covering ergonomics, workplace design, health and safety, task organization.* England: Wiley.

Ehret, Brian D. (2002). Learning where to look: Location learning in graphical user interfaces, *Proceedings of ACM INTERCHI 2002 Conference on Human Factors in Computing Systems*, 211-218.

Findlay, J.M. (1982). Global visual processing for saccadic eye movements. *Vision Research*, 22(8), 1033-1045.

Fleetwood, M. D. & Byrne, M. D. (2002) Modeling icon search in ACT-R/PM. *Cognitive Systems Research, 3,* 25-33

Gould, J. and Dill, A. (1969). Eye-movement parameters and pattern discrimination. *Perception and Psychophysics*, 6, 311-320.

Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye Movements and Visual cognition: Scene Perception and Reading*. New York: Springer-Verlag.

Kieras, D.E., & Meyer, D.E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic Mechanisms. *Psychological Review*, 104, 3-65.

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.) *Eye Movement Research: Mechanisms, Processes, and Applications.* New York: Elsevier Science Publishing.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1*(4), 201-220.

Tullis, T. S. (1997). Screen Design. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.) *Handbook of Human-Computer Interaction*, 2nd Edition. New York: Elsevier Science Publishing.

Williams, L.G. (1967). The effect of target specification on objects fixated during visual search. *Attention and Performance 1*, 355-360.

Zelinsky, G.J., and Sheinberg, D.L. (1997) Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1), 244-262.