Modeling Visual Search of Displays of Many Objects: The Role of Differential Acuity and Fixation Memory

David Kieras (kieras@umich.edu)
Electrical Engineering & Computer Science Department, University of Michigan
2260 Hayward Street, Ann Arbor, Michigan 48109 USA

Abstract

This paper describes a classic data set on visual search of 100-object displays that differ in size, shape, and color and presents a cognitive architecture model based on the active vision concept that accounts for the effects using differential visual acuity and fixation memory provided by a persistent visual store. The results provide an approximate upper bound on the duration of fixation memory, and some approximate acuity functions for modeling visual search.

Keywords: visual search; cognitive modeling; eye movements.

Introduction

Many everyday and work activities involve visual search, the process of visually scanning or inspecting the environment to locate an object of interest that will then be the target of further activity. An especially tractable form of visual search takes place in many human-computer interaction tasks in which a particular icon coded by color, shape, and other attributes must be located on a screen and then clicked on using a mouse. Such visual search takes place in a visual environment that is much simpler than natural scenes, and so is a both a good theoretical and practical domain to model visual search processes: it combines relative simplicity of the visual characteristics of the searched-for objects with practical relevance: the task is a natural one in the sense that such activities are very common in current technology; an example is current radar displays in military applications, which can contain a large number of icons and other objects (cf. Kieras & Marshall, 2006). Thus understanding in detail how visual search works in such domains can lead to better system designs.

This paper presents a model for the results of a classic study on visual search of large and dense displays of multiple items that can be searched by multiple attributes. This paper follows Kieras (2009), who presented a model for the Peterson et al. (2001) results demonstrating memory for fixations in a visual search task. In the Peterson et al. task, a single object, identified by shape, had to be located in field of a dozen objects which were very small and widely separated, meaning that each object had to be fixated before it could be identified. This paper presents a model for a task at the other extremes: A large number of objects, differing in several attributes had to be searched, but they were large enough and closely spaced enough that the properties of several objects could be considered in a single fixation. Memory for fixations still plays a role, but a critical role is also played by the differential availabilities of visual properties in extra-foveal vision, termed *differential acuity* in what follows.

Visual Search and Active Vision

In a laboratory visual search task, a display of objects is presented, and the participant must locate a particular object specified by perceptual properties and make a response based on whether such an object is present or exactly which properties it has (e.g. the specific shape). In most experiments, the display is static and contains some number of objects, only one of which is the target that must be responded to; the others are distractors. The properties of the display or the displayed objects are manipulated, and reaction time (RT) and/or eye movements are measured.

The empirical literature on visual search was dominated for a long time by studies that measured only RT, and often for tachistoscopically presented displays that ruled out eye movements. But more recently the cost of eye movement data collection has decreased to the point that it has become much more common, and deservedly so. While any behavioral measurement only indirectly reflects the mental processes that produce it, RT is clearly much less diagnostic of what goes on during visual search than eye movements. Furthermore, tasks in which the eye is free to move about a static display in a naturalistic manner, typical of eye movement studies of visual search, will be more representative of the normal operation of the visual system and the role of attention in visual activity. This point was argued eloquently by Findlay & Gilchrist (2003) in presenting an *active vision* framework for understanding visual activity; it is markedly different from traditional approaches to visual attention which have ignored both the role of eye movements and extra-foveal information.

In active vision, a key process is choosing the next object for inspection. A variety of studies (see Findlay & Gilchrist, 2003, for a review) have shown that this choice is not at all random; rather the color, shape, size, orientation, or other visual properties of objects influences which object is chosen for the next fixation; the phenomenon is called *visual guidance*. In the active vision framework, these properties are available in extra-foveal or peripheral vision to some extent, meaning that visual attention, which in the context of normal visual activity is almost synonymous with where the eye is fixated, is a process of selecting for detailed examination one of a large number of objects currently perceived to be in the visual scene, and doing this selection on the basis of the visual properties available in extra-foveal vision.

The availability of a perceptual property in extra-foveal vision depends heavily on the eccentricity (the distance in degrees of visual angle from the center of gaze) of the object, normally referred to in degrees of visual angle, and on the size of the object (also measured in degrees of visual angle), and on the specific property involved. For example, the color of an object of a given size in the periphery is usually more likely to be visible than its shape.

The EPIC Cognitive Architecture

The EPIC architecture for human cognition and performance directly supports an active vision approach to visual search and provides a general framework for simulating a human interacting with an environment to accomplish a task. Due to lack of space, the reader is referred to Kieras (2004), Kieras & Meyer (1997), Meyer & Kieras (1997) for a more complete description of EPIC.

The EPIC architecture consists of software modules for the simulated task environment, or device, that interacts with a simulated human, which consists of perceptual and motor processor peripherals surrounding a cognitive processor. The device and all of the processors run in parallel with each other. To model human performance of a task, the cognitive processor is programmed with production rules that implement a strategy for performing the task. When the simulation is run, the architecture generates the specific sequence of perceptual, cognitive, and motor events required to perform the task, within the constraints determined by the architecture components and the task environment.

Figure 1 shows the visual system of EPIC. The *eye processor* explicitly represents differential retinal availability in terms of *acuity functions* that specify whether each visual property of each object is currently visible as a function of the size of the object and its eccentricity. The currently available visual properties for each object are represented in the *sensory store*; the *perceptual processor* then encodes the properties of each object, possibly in relation to other objects, and passes the encoded representation on to the *perceptual store* where they are available to the cognitive processor to match the conditions of production rules. The perceptual store thus contains the current representation of the visual world that cognition can reason and make decisions about, including decisions about where to move the eyes next by commanding the *ocular motor processor*. The perceptual store retains the representations for *all objects currently visible*, with more information and detail about those that have been fixated.

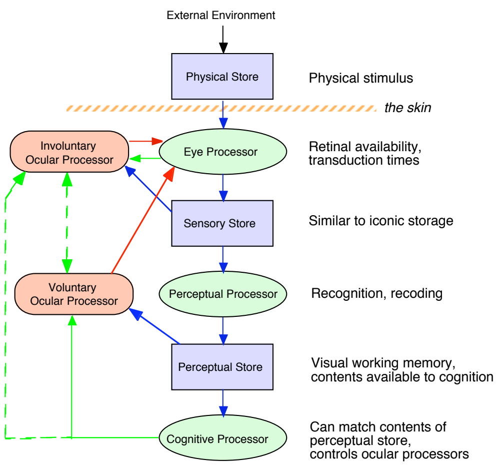

Figure 1. EPIC's visual system.

Persistence of the visual perceptual store

When the eyes move away from an object, the properties of the object persist for a short time (e.g. 200 ms) in the sensory store, and when lost, the perceptual processor notes that the corresponding property in the perceptual store no longer has sensory support. After a relatively long time, the property will then be lost from the perceptual store. But if the object disappears completely, it and all of its properties will be removed from the perceptual store fairly quickly.

The concept is that as the eyes move around the visual scene, a complete and continuous representation of the objects currently present in the visual situation will be built up and maintained in the perceptual store, allowing the cognitive processor to make decisions based on far more than the properties of the currently fixated object. The notion that this information persists for a considerable time as long as the scene is present is supported by studies summarized by Henderson & Castelhano (2005): a naturalistic visual scene is continuously present, but using a gaze-contingent forced-choice paradigm, subjects are tested for their memory of a previously fixated object; retention times at least several seconds long were observed. The model for the Peterson task (Kieras, 2009) provided a good fit to the repeat-fixation data with a retention time of at least 4 sec.

The Williams Study

A classic study using early eye-movement recording methodology was done by Williams (1966, 1967), who ventured into experimental territory commonly avoided even today. This study manipulated the size of the objects along with their color and shape, an unusual combination in the visual search literature, and used a very large number of objects, which provides an upper bound on the difficulty of search tasks of this sort.

The task required visual search of 100 objects varying in size, color, and shape, each with a unique two-digit label. The 100 objects represented all combinations of 4 sizes, 5 colors, and 5 shapes. The search task was to locate the object with the matching label. Depending on the experimental condition, additional attributes of the target object were cued; all combinations of size, color, and shape cues were tested in addition to the *Number-only* cue, which was only the object label. The hypothesis was that if a specification is an effective cue for visual guidance, more fixations should be on objects matching the cue than expected by chance.

The entire display is 39° X 39° (all degrees are degrees of visual angle), and the search objects range from 0.8° to 2.8° in size and distributed at random into the 13 X 13 grid of 3° X 3° cells. The cue specifications were shown in the center of the display. To convey an overall impression of the task, Figure 2 provides an example display produced by the model to be described later. Due to space restrictions this figure is too small for the details to be visible in a paper printing, especially in monochrome, but the details can be seen easily by zooming in with the original pdf file. In this example, the specified target is the medium-size

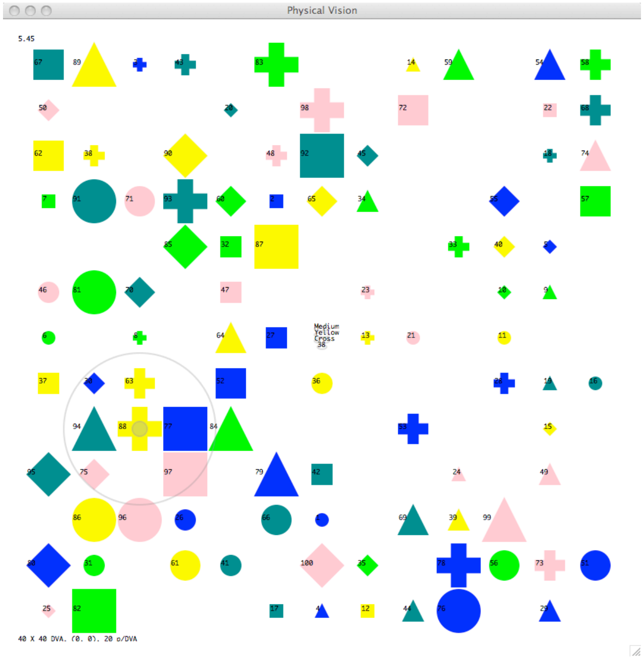

Figure 2. An example of the physical display in a Williams (1966) task trial after several fixations as depicted in EPIC's automatically-generated display. Zoom in on this figure in the pdf file to see the detail.

yellow cross labeled 38, which is in the upper-left of the display. The concentric circles at center left show the current location of EPIC's eyes; the small inner circle has a 1° diameter corresponding to the conventional fovea size; the outer circle is a calibration ring with 10° diameter. The display is shown to scale, except that to maintain legibility, the numeric labels are shown as magnified and left-justified in the object bounding boxes; in the actual stimuli and model representation, they are only 0.3° high, which would require foveation to recognize, and centered in the object.

The specification names for color and shape were the obvious names, but the four sizes were described as *small*, *medium*, *large*, and *very large*. The specifications appeared first in the center of the display; when a button is pressed, the search objects were added to the display. The participant pressed another button when he or she had located the specified object.

Eye movements were recorded with a corneal-reflection film camera system and scored by hand. The total number of fixations were counted, and classified by whether they fell on objects whose size, color, and shape matched the specifications. While 61% of the fixations were attributed to a specific object, 29% were deemed unclassifiable, a relatively large number by current methodological standards.

Unlike modern practice, Williams obtained approximate reaction times (RT) *indirectly* by counting the number of fixations and dividing by 3.25, the observed average number of fixations per second. Because the observed number of fixations and the reported RTs are perfectly correlated, the RTs will only be mentioned occasionally.

The Data

This being an early and basically descriptive study, Williams did not report confidence intervals or information sufficient for their calculation, and conventional statistical tests were not relevant. However, the data set consisted of many thousands of fixations collected from 30 participants who performed 200 trials spread over 8 conditions. Based on the original reports, it appears that a typical sample size for the statistics for any one condition as reported below is in the neighborhood of about 1000. The proportions of fixations on objects of various types are the most important results; for an observed proportion of 0.5, the 95% binomial confidence interval for a sample size of 1000 is about 0.47-0.53; this ±0.03 range can be used as an approximate confidence interval for this important subset of the data.

Figure 3 shows the observed proportion of fixations on objects that matched the cued properties (the predicted values will be discussed below). E.g., if the color was the only specified cue, about 60% of the fixations were on objects with the specified color. Figure 4 shows the number of fixations for each cue type.

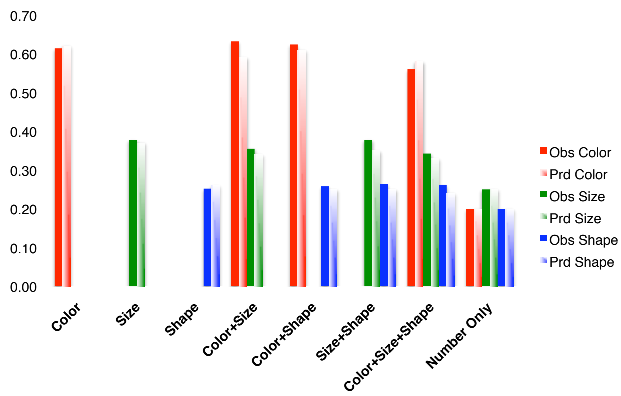

Figure 3. Observed (solid bars) and predicted (shaded bars) proportion of fixations on objects that matched each cue type. The 95% confidence intervals would be roughly ±0.03 for each observed proportion.

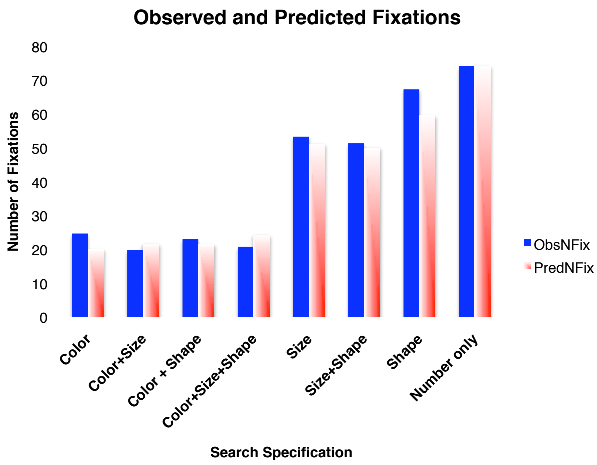

Figure 4. Observed and predicted number of fixations for each cue type.

Visual guidance produced by color, size, and shape

It is clear from the results that color is the strongest cue for visual guidance, resulting in the highest proportion of fixations on matching objects (0.61), the fewest fixations (25) and the fastest RTs (not shown, 7.6 s). Size comes next, and shape is a distant third. There is a tendency for each cue to have little or no effect if a stronger cue is also present. If only the label is provided (the *Number-only* cue), the fixations on objects that match the target properties is at chance level, the number of fixations is large (74), and the RT is quite long (23 s).

The importance of color in visual search is consistent with many results ranging from classic human factors studies (e.g. Sanders & McCormick, 1987) to recent HCI-oriented studies (e.g. Fleetwood & Byrne, 2006). But in the active vision framework, color is not specially privileged in some way, but rather, various direct measurements show that the color of an object is visible over a wide range of eccentricity and object sizes (e.g. Gordon & Abramov, 1977), and so can often serve as an effective cue about where to look next. The relative ineffectiveness of shape is likewise not due to a fundamental problem with shape, but rather that in many cases, recognizing the shape requires resolving detailed features that can only be seen close to the fovea. As an extreme of shape recognition, recognizing the text label involves detecting small features, and so requires foveation unless the text is quite large (Anstis, 1974).

Repeat fixations and memory failures

One overall feature of these results is that many more fixations are required than should be necessary if each object only received one fixation; for example, it should require no more than 50 fixations on average in the Number-only condition to find the labeled object. Williams reports a small number (3%) of immediate repeat fixations, but does not report how many repeat fixations appeared over longer time periods. Apparently objects are frequently looked at repeatedly; e.g. the 74 fixations in the Number-only condition implies a repeat rate of about 33%!

In contrast, recent observation and modeling of repeat fixations (see Peterson et al. 2001, Kieras & Marshall, 2006, Kieras, 2009) suggests that repeat fixations are relatively rare, around 5%, implying a good memory for previous fixations, and almost all are performed immediately, being due to recognition (encoding) failures rather than failures of the memory for previous fixations. The 3% immediate repeat rate reported by Williams is consistent with this, but not the much higher overall repeat rate implied by the total number of fixations.

However, the low-rate results were obtained in search tasks involving many fewer objects and that took much less time than Williams' task. Perhaps the much higher repeat rate in Williams' results is due to time decay of the fixation memory. In fact, in Peterson's task, repeat fixations at long lags become more frequent if the trial has gone on for an unusually long time (Peterson, personal communication). This issue will be important in modeling the Williams data.

Model for the Williams Task

Constructing an EPIC model for the Williams task required a choice of (1) visual acuity parameters, (2) a parameter for the decay time of visual properties in the perceptual store that are no longer sensorily supported, and (3) a set of production rules that implemented the visual search strategy. Each of these will be described briefly.

Acuity functions

Despite the many decades of research on vision, the literature does not contain a comprehensive set of parametric data on acuity for different visual properties as a function of their eccentricity and size, especially for the properties and values typical of computer displays. Space limitations do not allow even a cursory review of the available data (but see Findlay & Gilchrist, 2003). To deal with this non-definitive picture, a simple family of acuity functions were proposed, and their parameters determined by a combination of general constraints set by the literature and iterative maximization of fit in the models. A separate function was specified for each property: encoded size (*small*, *medium*, etc.), color, shape, and text label. The text acuity function was specified as text being available within 1° of the current eye position, corresponding to the conventional definition of foveal vision and the small size of text used. A psychophysical acuity function was used for the other properties: For the property to be available, its size s must exceed a threshold which increases quadratically with eccentricity e and includes a Gaussian noise component X whose variability increases with the object size and coefficient of variation v:

$$threshold = ae^2 + be + c$$
$$P(available) = P(s + X > threshold)$$
$$X \sim N(0, vs)$$

Such a function produces a wide area of highly probable availability, with a sharp tapering-off towards the periphery. The quadratic form was selected for simplicity: the parameters can be easily set to reflect a minimum size, general trend, and degree of curvilinearity, and were set to be consistent with models for other tasks not described here, and to have as much uniformity in the parameter values as possible. The function for color availability used in the model had parameter values of $v=0.7$, $a=0.035$, $b=0.1$, $c=0.1$. The acuity functions for encoded size and shape had the same values except for larger quadratic coefficients a of 0.2 and 0.3 respectively. Thus, consistent with the literature, the availability of the size and shape properties drops off with eccentricity much more rapidly than for color.

The availability for each property at the retinal and sensory store level is independently resampled for all objects whenever the eye is moved. Figure 5 shows an example of EPIC's visual sensory store after several fixations, corresponding to Figure 2, showing what is currently available around the fixation point. In EPIC's display, objects whose location, but no other properties, are known are represented as light gray open circles. Objects which are close enough to the current fixation point to have their color available, but not their shape, are represented as colored open circles. In Figure 5, the shape, color, encoded size, and label are available for the currently fixated object. The colors of several extrafoveal objects are also available, and even the shape for a nearby large object. As the eye moves around, the available properties of the same object

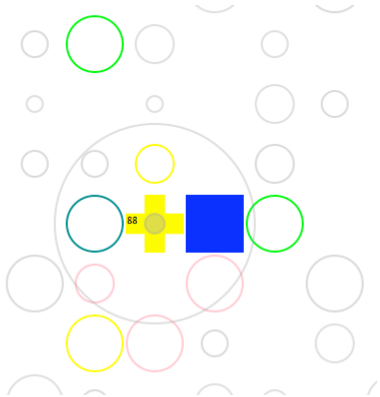

Figure 5. An example of the contents of the sensory store corresponding to the lower left corner of Figure 2, showing available properties of objects near the current fixation point.

can fluctuate, and will not be reliably available from one fixation to the next.

Perceptual store persistence time

Once a property of an object is visible, that property is attached to the object representation in the visual perceptual store where it can serve to match conditions of production rules. The visual perceptual store is persistent, in that as long as an object is within the visual field, its properties, once acquired, will persist for a long time and thus can serve as a memory for previous fixations, as described in Kieras (2009). Figure 6 shows a sample of EPIC's visual perceptual store, corresponding to Figures 2 and 5, several seconds into the visual search, showing the information persisting from previous fixations. Previously fixated objects have all properties including the label, but will eventually lose this information until fixated again. But in the meantime, their color, size, or shape can be used to guide the choice of which object to fixate next.

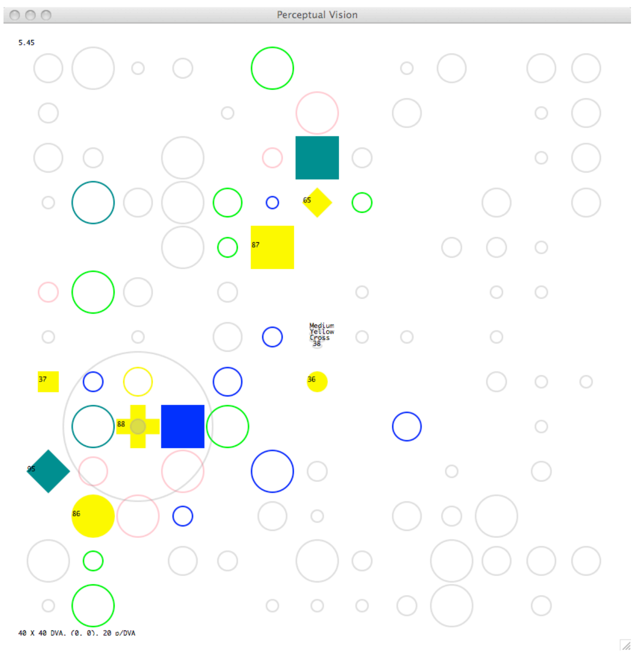

Figure 6. An example of the contents of the perceptual store after several fixations, showing the accumulated object information. Zoom in on this figure in the pdf file to see the detail.

The duration parameter was estimated iteratively by fitting the model, starting with the 4 sec lower bound determined in Kieras (2009); a good fit was found with a duration of 9 seconds.

Task strategy

The visual search strategy in the model is an application of a basic strategy, shown in Figure 8, that has been used in several EPIC visual search models. There are two threads of execution. Nomination rules in the first thread propose objects to fixate based on available visual properties, and also nominate a random choice. Choice rules then pick a single candidate from the nominated objects according to a priority scheme, and launch an eye movement to the chosen candidate. The rules in the second thread wait for all relevant properties of the fixated candidate to be fully visible and either respond if it is a target, or discard the candidate if not. Given the typical 100 ms transduction and encoding times for visual properties and the 50 ms production rule cycle time, the overlapped processing provided by the two threads enables the time between successive eye movement initiations to be short, in the range of 250 to 300 ms, which is commonly observed in high-speed visual search tasks.

For the Williams model, the strategy nominates candidate objects that have the cued properties, such as the cued color or cued shape. The fixation memory effect is implemented by only nominating objects whose text label property is currently unknown; either because the object was never fixated, or it was fixated a long time ago and has been lost from the perceptual store. The priority scheme for choosing a fixation target favors the more available information, and so chooses an object with a matching color over one with a matching size over one with a matching shape. For simplicity, given the apparent very high repeat fixation rates in the data, the mechanism for the relatively rare

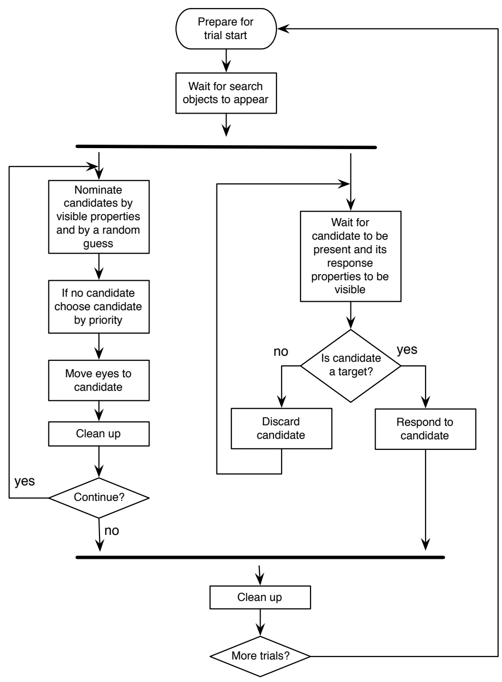

Figure 8. Flowchart for the search task strategy.

encoding failures used in previous models (e.g. Kieras, 2009; Kieras & Marshall, 2006) to trigger repeat fixations was not implemented in this model, corresponding to an assumption that most of the revisits are due to memory failure in this task.

Model Results

The model was run for 500 trials in each experimental condition, and the simulated eye movements and response time data were collected and tabulated analogously to the original experiment. Figure 3 above shows the observed and predicted proportion of fixations of each type. Clearly the fit is very good using the acuity function and perceptual store persistence parameters listed above (R^2 = .99; average absolute error (AAE) = 3%).

The observed and predicted number of fixations is shown in Figure 4 above. Again there is a very good fit (R^2 = 0.98, AAE = 12%). The observed and predicted RTs (not shown) also fit well (R^2 = 0.98 and AAE = 9%), although there is a general tendency for the model RTs to run longer than William's results. Given the unusual methodology used to determine the RTs, it is not clear that attempting to improve the fit to the absolute value would be worthwhile.

In an analysis of the model output, the proportion of repeat fixations was found to increase substantially as the perceptual store duration was decreased, and the number of fixations (or RT) increased. The persistence parameter was adjusted to produce the overall good fit on the number of fixations shown in Figure 4, and the proportion of repeat fixations on search objects was then determined with the final parameter value. The range was 11% repeats in the best condition to 33% in the Number-only condition. This proportion was highly linear with the predicted number of fixations (R^2 = 0.95). This suggests that the loss of fixation memory over time is a good account for the excess number of fixations in the data.

Conclusion

This model, along with the one in Kieras (2009), represents a realization of the active vision concept in terms of a computational cognitive architecture that incorporates differential acuity and a persistent visual store that represents the current visual situation and provides a memory of previous fixations. Two more specific points emerge: (1) Simplistic statements about which properties can guide visual search must be replaced by statements about which properties are available in a specific visual situation. For example, color should not be very effective if the objects were very small, and shape should be more effective if the objects were larger. (2) Repeat fixations have two causes: the persistent visual store is capacious and reliable at short durations, meaning that repeat fixations are due just to encoding errors, but if the search takes a very long time, information from previous fixations is lost, and more repeat fixations are the result.

This general model appears to be ready for practical application in situations where the to-be-searched display contains uniform-color objects with simple geometric shapes and very small distinguishing features such as text labels. The specific acuity functions determined here should be useful approximations in modeling such displays.

At the theoretical level, this type of model appears to be a simple and sound approach to representing visual activity, and is ready to use either as a component in models of more complex tasks that involve visual search as a subtask, or as a basis for models of more advanced visual processing.

Acknowledgment

This work was supported by the Office of Naval Research, under Grant No. N00014-06-1-0034.

References

Anstis, S.M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision research*, 14, 589-592.

Findlay, J.M., & Gilchrist, I.D. (2003). *Active Vision.* Oxford: Oxford University Press.

Fleetwood, M. D. & Byrne, M. D. (2006). Modeling the Visual Search of Displays: A Revised ACT-R Model of Icon Search Based on Eye Tracking Data. *Human Computer Interaction*, 21, 153-197.

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. Hue and saturation. *Journal of the Optical Society of America*, 67(2), 202-207.

Henderson, J.M. & Castelhano, M.S. (2005). Eye movements and visual memory for scenes. In G. Underwood (Ed.), *Cognitive processes in eye guidance*. New York: Oxford University Press. 213-235.

Kieras, D.E. (2004). EPIC Architecture Principles of Operation. Web publication available at ftp://www.eecs.umich.edu/people/kieras/EPIC/EPICPrinOp.pdf

Kieras, D. (2009). The persistent visual store as the locus of fixation memory in visual search tasks. In A. Howes, D. Peebles, R. Cooper (Eds.), *9th International Conference on Cognitive Modeling – ICCM2009*, Manchester, UK

Kieras, D.E, & Marshall, S.P. (2006). Visual Availability and Fixation Memory in Modeling Visual Search using the EPIC Architecture. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 423-428.

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

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

Peterson, M.S., Kramer, A.F., Ranxiao, F.W., Irwin, D.E., & McCarley, J.S. (2001). Visual search has memory. *Psychological Science*, **12**, 287-292).

Sanders, M. S., & McCormick, E. J. (1987). *Human factors in engineering and design* (6th ed.). New York: McGraw-Hill.

Williams, L.G. (1966). A study of visual search using eye movement recordings. Technical Report, Honeywell Inc., Feb. 28, 1966. NTIS AD629624.

Williams, L.G. (1967). The effects of target specification on objects fixated during visual search. In A.F. Sanders (Ed.) *Attention and Performance*, North-Holland. 355-360.