Human Gaze Control in Real World Search

Daniel A. Gajewski[1,4], Aaron M. Pearson[1,4], Michael L. Mack[2,4], Francis N. Bartlett III[3,4], and John M. Henderson[1,4]

[1] Michigan State University, Department of Psychology, East Lansing MI 48824, USA
{dan, aaron, john}@eyelab.msu.edu
http://eyelab.msu.edu
[2] Michigan State University, Department of Computer Science, East Lansing MI 48824, USA
mike@eyelab.msu.edu
[3] Michigan State University, Department of Zoology, East Lansing MI 48824, USA
bartle47@msu.edu
[4] Michigan State University, Cognitive Science Program, East Lansing MI 48824, USA

Abstract. An understanding of gaze control requires knowledge of the basic properties of eye movements during scene viewing. Because most of what we know about eye movement behavior is based on the viewing of images on computer screens, it is important to determine whether viewing in this setting generalizes to the viewing of real-world environments. Our objectives were to characterize eye movement behavior in the real world using head-mounted eyetracking technology and to illustrate the need for and development of automated analytic methods. Eye movements were monitored while participants searched for and counted coffee cups positioned within a cluttered office scene. Saccades were longer than typically observed using static displays, but fixation durations appear to generalize across viewing situations. Participants also made longer saccades to cups when a pictorial example of the target was provided in advance, suggesting a modulation of the perceptual span in accordance with the amount of information provided.

1 Introduction

The most highly resolved portion of one's visual field is derived by the foveal region of the retina where the cones are most densely packed. Because this region of high acuity is very small, corresponding only to about two degrees of visual angle, the eyes must be directed toward points of interest in a scene in order to encode visual details. Gaze shifts tend to occur at a rate of around three to four times per second with visual information extracted from the environment primarily when the direction of one's gaze is relatively stable. These periods of stability, called fixations, are separated by rapid movements of the eye, called saccades. A fundamental concern for those studying visual perception is the

L. Paletta et al. (Eds.): WAPCV 2004, LNCS 3368, pp. 83–99, 2005.

control of eye movements during scene viewing: What are the processes that determine the time spent engaged in fixations and govern the selection of targets for upcoming saccades?

Eye movement control is of interest to cognitive psychologists because eye movements are overt manifestations of the dynamic allocation of attention. While attention can be directed covertly away from the center of fixation, the natural tendency is for eye movements and attention to remain coupled. The most prevalent view of the relation between attention and eye movements is one where eye movements are preceded by shifts of attention to the location toward which the eyes are moving [1],[2],[3],[4],[5]. Sequential attention models such as the one proposed by Henderson [1] posit a relation between attention and eye movements that begins with attention allocated to the foveated stimulus. When processing at the center of fixation is complete or nearly complete, attention is disengaged and reallocated to a more peripheral location. This reallocation of attention coincides with the programming of an eye movement that brings the center of vision to the newly attended region of the visual field. In this view, questions of eye movement control are really questions of attentional selection because attentional selection and the targeting of saccadic eye movements are functionally equivalent.

Eye movement control is additionally of interest because of the degree of intelligence reflected in eye movement behavior. Early studies of picture viewing, for example, revealed a tendency for fixations to cluster in regions that were considered interesting and informative [6],[7],[8], (see [10] and [11] for reviews). Additionally, empty regions frequently did not receive fixation, suggesting that uninformative regions could be rejected as saccade targets on the basis of peripheral information. Subsequent research has been aimed at determining the roles of both visual and semantic informativeness in gaze control. A number of recent studies demonstrate a correlation between fixation densities and local image properties that can be represented in a wide range of statistics. For example, fixations tend to fall in regions that are high in local contrast and edge density compared to regions that do not receive fixation [12],[13],[14]. While subsequent research suggests that fixations are not drawn by local semantic information early in scene viewing [15],[16], (but see [17]), global properties of a scene, such as its meaning and overall spatial layout, can exert an influence as early as the first fixation. For example, participants in a search task were able to find target objects in scenes more quickly when a brief preview of the scene was presented in advance [18]. Finally, viewing patterns are influenced by the cognitive goal of the observer. In a classic study, Yarbus [9] monitored eye movements during the presentation of a picture of Repin's *An Unexpected Visitor*. The observed gaze patterns were modulated by the question posed to the viewer in advance, suggesting that the informativeness of scene regions is task-dependent.

While there have been numerous advances in the domain of eye movement control in scenes, most of what we know is derived from relatively artificial viewing situations. Eye movements are typically recorded while research participants view images presented on a computer monitor. The most commonly used eye

tracking systems are considered stationary because the viewer's head position is maintained using an apparatus that includes a chin and forehead rest. Recent innovations in eyetracking technology, however, allow for the monitoring of eye movements as participants view real-world environments [19]. With these head-mounted eyetracking systems, eye movements are recorded along with a video image of the person's field of view provided by a camera positioned just above the eyes. Head-mounted eyetracking systems are advantageous because they allow researchers to study eye movements in more ecologically valid settings. For example, the field of view provided by a computer monitor is rather limited, as the display only extends to around 20° of visual angle. Head-mounted eyetrackers, on the other hand, allow eye movements to be monitored with a field of view that is unconstrained. In addition, allowing the head to move freely provides a more naturalistic viewing situation and allows for the study of eye movements during tasks that require interactions with the environment.

In the present study we wished to begin to address the question: How well does viewing on computer displays generalize to the viewing of natural environments? To investigate this question, we set up a simple search task in a real-world environment. Participants were asked to count coffee cups that were placed in various locations in a professor's office. The task provided the opportunity to examine a number of basic properties of eye movement behavior in a naturalistic setting.

First, we wished to determine the saccade lengths to objects of interest. Studies using static scenes on computer displays have predominately found mean saccade amplitudes in the $3-4°$ range [15],[16],[10]. Loftus and Mackworth [17], however, observed saccade lengths to objects that were as high as 7 degrees. Henderson and Hollingworth [10] suggested that this anomalous finding could reflect differences in stimuli. The Loftus and Mackworth [17] stimuli were line drawings of scenes that were relatively sparse compared to those used in other studies. The reduction in contours with stimuli of this kind would be expected to decrease the amount of lateral masking thereby increasing the distance from the center of vision from which useful information can be acquired (i.e., the perceptual span). While scene complexity was not manipulated in the present study, one would expect to find small saccades to cups because the room scene that was employed was cluttered with objects. Alternatively, saccade lengths to target objects might not generalize across viewing situations. In this case, saccade lengths could be longer in the natural environment than observed in previous studies using computer displays.

Second, because a complete understanding of eye movement control in real-world scenes will require a general knowledge of the basic properties of how the eyes move through a scene, we wished to generate fixation duration and saccade length distributions for the entire viewing episode. Mean fixation durations in search tasks have been reported as low as 275 ms [20] and 247 ms [16]; modal fixation durations around 220 ms have been reported in a search task using line drawings of scenes [16]. Mean and modal saccade lengths in picture viewing tasks have been reported as low as 0.5° and 2.4° respectively [10]. Recent data, how-

ever, suggest that smaller fixation durations and longer saccades may be more common in dynamic and naturalistic viewing situations [21]. Indeed, because the range of possible saccades in static displays is severely limited, determining saccade lengths in natural viewing situations is central to the question of the generalization of eye movement behavior.

Third, in addition to exploring the issue of generalizing extant eye movement data to the viewing of natural environments, we wished to begin exploring the concept of the perceptual span in real-world scenes. While the perceptual span has been extensively studied in the context of reading [20], very little has been done using scenes as stimuli; and to our knowledge there is nothing in the literature that addresses the perceptual span during the viewing of natural environments. In the present study, we wished to determine whether the perceptual span could be influenced by the precision of the representation of the target object. To investigate these issues, we compared the saccade lengths to two different sets of target items. One group of participants searched for coffee cups that came from a matched set and a second group searched for a mixture of coffee cups that varied in color, shape, and design. Importantly, the group that searched for matched cups were shown a picture of the target cup before the search began. If having a more precise representation of the target increases perceptual span, saccades to cups should be longer in the matched cup condition. This would be expected because participants in the matched cup condition would have more information available with which to reject regions as uninformative and select regions that are likely to have a cup.

Finally, while the present research was designed primarily to investigate a number of basic properties of eye movement control during the viewing of natural environments, it also represents the development of analytic tools for head-mounted eyetracking data. Despite the gains associated with using head-mounted eyetrackers, the utility of this technology is limited because data analyses are tedious and time-consuming. The greatest challenge is driven by the fact that the data are recorded in a reference frame that is dissociated from the world that the participant is viewing. With stationary eyetrackers, fixations are given in image plane coordinates that correspond to where the participant is looking on the computer display. The mapping from eyetracker space to display space is straightforward because the image plane and real-world coordinate systems coincide. Analyses can be automated because the locations of the fixations in the image plane coordinate system always correspond to the same locations on the display. Fixations are also given in image plane coordinates with head-mounted eyetrackers, but these coordinates correspond to varying locations in the real world because the image reflects a changing point of view. As a result, data analyses are often based on the hand scoring of video images that include a moving fixation cross indicating where the viewer is looking. The locations of fixations are manually determined by stepping one-by-one through the frames of video. This approach is extremely labor-intensive. Given that there are thirty frames for every second of viewing with a 30 Hz system (and more with faster systems), and every saccade and fixation must be determined, generating overall

saccade length and fixation duration distributions with hand-scoring methods is usually impractical if not impossible. In this paper we report a method of automatically extracting the eyetracking data and we contrast the results from this method with the results derived from the hand-scoring method.

2 Present Research

2.1 Methods

Participants - Twenty-six Michigan State University undergraduates participated in exchange for course credit. All had normal or corrected-to-normal vision.

Stimuli - The room scene was a professor's office, approximately 8 x 12 meters, containing typical office items such as a desk, computer, chairs, file cabinets, and various shelving filled with books and paperwork (see Figure 1). The cups were placed in six locations throughout the room, spaced at relatively uniform distances from one another. The six cups used in the matched condition were all exactly the same. The other six cups were a variety of colors, shapes, and sizes.

Apparatus - Participants wore an ISCAN model ETL-500 head mounted eyetracker. This eyetracker consists of a pair of cameras securely fastened to a visor that is strapped on the participant's head. One camera records the scene that the participant is currently viewing. The other, a camera sensitive to infrared light, records the movement of the participant's left eye. An infrared emitter that is housed within this camera illuminates the eye. Because this emitter is secured to the participant's head, the corneal reflection stays relatively stable across head and eye movements as the pupil moves. By measuring the offset between the pupil and the corneal reflection, it is possible to identify the location in the scene images that the participant is fixating. Since the visor is enclosed, participants were able to view 103° of visual angle horizontally, and 64° vertically. A plastic shield on the visor that is used to block infrared radiation when tracking the eyes outside was removed for this study, as it affects color vision.

In the ISCAN system, the scene and eye video are merged by a video multiplexer that takes input from both cameras simultaneously at 30 Hz. The multiplexed video, a composite video with the eye image on the left and scene image on the right, is then recorded by a Mini DV recorder. The final output is a scene video with a small cross on the image that corresponds to the center of fixation. The location of this cross is accurate to approximately one degree of visual angle under optimal conditions.

Procedure - Participants were tested individually and all testing was accomplished in one day to maintain control of the search environment. The session began in a laboratory adjacent to the professor's office. Upon arrival, participants were briefed about the experiment and equipment. The participants in the matched condition were additionally shown a pictorial example of the specific type of coffee cup they were to search for and count. Participants were then

fitted with the eyetracker and calibrated. The calibration involved sequentially fixating five points arranged on a blackboard approximately 6 meters away. After calibration, they were walked to the professor's office. Before entering the office, the task instructions were repeated. Participants were then told to close their eyes and were led into the office by an experimenter. Upon entering the office, an experimenter placed the participants' feet over two pre-designated marks on the floor. This was done to ensure that each participant viewed the room from the same position. Once the participants were properly positioned, the office light was switched on and they were told to open their eyes and begin counting coffee cups out loud. They were given 15 seconds to count as many cups as they could find. Participants were aware of the limited amount of search time and were thus encouraged to find the cups as quickly as possible. When the time expired, an experimenter shut off the lights and told the participant to close their eyes. The participant was then led back to the laboratory for equipment removal and a memory test was administered to address questions that were not central to the topic of this paper. Once the memory test had been completed, participants were debriefed and allowed to leave.

2.2 Analysis

The objectives of the present study required two analytic methods. The first of these consisted of a hand-scoring method to determine saccade distances to each cup as well as the duration of the first fixation on each cup. The second consisted of developing and implementing an automated scoring method to eliminate subjectivity in separating fixations from saccades as well as for measuring the distance traveled during saccades and the durations of individual fixations. The hand-scored data were used as a benchmark to assist in developing the automated method of data analysis as well as in conjunction with the automatic analyses in order to converge on a common solution.

Before any scoring could begin, a number of intermediary steps were necessary to convert the video data into a workable format. The ISCAN ETL-500 data, in its raw form, is recorded onto 8mm digital videotape. At this point, the data are multiplexed and recorded onto a single tape to synchronize the separate video streams captured from the head-mounted eyetracker (HMET) eye camera and the scene video captured from the world camera.

The data are de-multiplexed into two video streams and recorded onto separate Mini DV (digital video) tapes via two Mini DV recorders. During this process, the ISCAN ETL-500 computer hardware and software interface utilize the pupil and corneal reflection from the eye video as well as the calibration sequence gathered during data collection to calibrate and display the point of regard of a given participant's eye onto the scene video (see Figure 1). This scene video is recorded onto DV tape along with the audio signal from the multiplexed Hi-8 video using one of the Mini DV recorders. The second recorder is used to record the video of the participant's eye.

Once the video stream has been separated and recorded onto DV tape, it is then transferred to computer via Adobe Premiere. The videos are then saved as

Fig. 1. An example of the scene video as it is output from the HMET hardware with a calibrated fixation cross indicating POR

avi files with a screen resolution of 720 x 480 pixels in their native DV format. This step maintains the original resolution of the DV videos as well as any compression associated with the DV format.

Scoring by Hand - Hand-scoring the video data involves locating the critical fixations within a series of digitized still images that were taken of the room scene while stepping frame-by-frame through the scene video output. The pixel distance from the launch fixation to the landing fixation was recorded for each cup that a participant counted. The pixel distances for the saccades to the cups were then expressed in terms of degrees of visual angle using a conversion factor determined for the still images. The conversion factor was derived by measuring the visual angle subtended by landmark features in the room (e.g., the length of shelves) from the participants' point of view. The visual angles subtended by the landmark features were then compared to the corresponding pixel distance in the still images to generate the conversion factor. Frame numbers for the onsets and offsets of the launching and landing fixations were also recorded to generate fixation durations for the fixations on the cups. Each video was scored by two experimenters and discrepancies were evaluated on a case-by-case basis.

Automated Scoring - Automated analysis of the eye and scene videos begins with syncing the videos to align corresponding frames. This is accomplished with an observer manually editing the videos so that the movements of the cross representing the point of regard (POR) in the scene video occur at the same frame number as the corresponding eye movements on the eye video.

Next, the center of mass of the pupil is recorded for each frame of the eye video. First, the pixels of the pupil region are extracted by converting the frames into binary images with a luminance threshold.

In many cases, the corneal reflection of the eye appears on the pupil. When this occurs, a poor extraction of the pupil region results from the threshold step

explained above. With the area of the pupil not fully extracted, the pupil's center of mass cannot be correctly determined. This problem is overcome by fitting an ellipse to the pupil region and analyzing the resulting region created by the fit. The following conic ellipse model is fit to the boundary pixels of the pupil region with a least squares criterion:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{1}$$

From this fit, the center of mass, area, and vertical diameter of the pupil are calculated and recorded (see Figure 2).

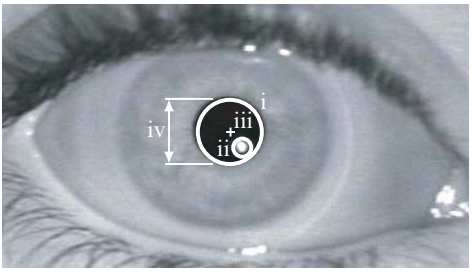

Fig. 2. Sample frame from eye video. i. Pupil ii. Corneal reflection iii. Center of mass of pupil iv. Vertical diameter of pupil

At this point, we determine when fixations, saccades, and blinks occur. A threshold is set for change in pupil area and change in pupil diameter to determine when a blink occurs. Fortunately, the change in pupil area and pupil diameter during a blink is much faster than during pupil dilation and constriction, so a threshold is easily set. Next, saccades are determined with a two-step process. Using the acceleration of the pupil's center of mass, a liberal decision of when a saccade occurs can be made with another threshold. By first using acceleration to find saccades, slower movements of the eye that keep the point of regard stable on the world while the head is in motion can be recognized and scored as fixations. The next step involves using a two-stage threshold for the velocity of the pupil's center of mass (see Figure 3). The first threshold finds the peak velocity of a saccade. Once this threshold has been exceeded, the frames before and after the current frame are progressively examined to find the tails of the saccade. This analysis results in each frame of the eye video labeled as part of a fixation, saccade, or blink.

Fixations are extracted from the frame labels by grouping together contiguous frames labeled as part of a fixation. Fixation durations are calculated by dividing the number of frames in a fixation by the frame rate of the eye video (e.g., a fixation with 5 frames recorded at 30 frames per second would yield a fixation duration of 167 ms). Saccade lengths are calculated by determining the Euclidian distance traveled by the center of mass in the eye video from one fixation to the next. This pixel distance is then converted to degrees of visual angle using

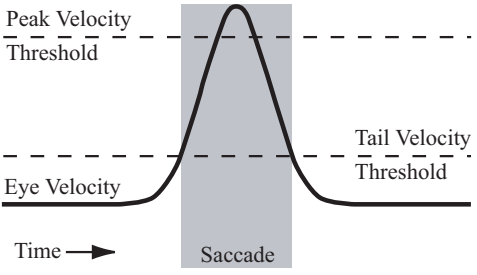

Fig. 3. A common velocity pattern during a saccade is shown by the black line. First, the middle of a saccade is found with a peak velocity threshold (top dotted line). Once this threshold has been exceeded, neighboring frames are progressively included in the saccade until the frames' velocities drop below the smaller tail velocity threshold (bottom dotted line). This algorithm allows for more accurate determination of the entire saccade

a conversion factor determined by measuring the distance traveled by the eye during calibration.

To locate the fixations corresponding to cups, the scene videos (as output from the HMET hardware) were divided evenly among three independent observers. Cup fixations were located utilizing the participants' audible responses to identify where the cup was counted, then stepping backward frame-by-frame from this point until a corresponding cup fixation was located in the video as indicated by the POR fixation cross. The frame number at which this fixation occurred was recorded into a spreadsheet. This allowed the frame locations at which cup fixations occurred to be noted in the fixation-listing file generated by the automated scoring method. The saccade lengths to the cups were the saccade lengths determined for the fixation on the cups in the fixation-listing file.

3 Results

Six (out of 26) participants were excluded from the analyses. One of them was removed due to an error during data collection resulting in data that was not analyzable. An additional participant was removed because it appeared that this person misunderstood the instructions. The other three participants were eliminated due to increased noise in the tracking caused by extraneous reflections in the eye image - making the data very difficult to analyze by hand, and impossible to cross-evaluate via the automated data collection software. All of the following analyses were based on the remaining 20 participants, 9 in the matched cup condition and 11 in the mixed cup condition.

3.1 Saccade Lengths and Fixation Durations for Cups

The mean saccade lengths to the cups were 10.7° and 11.9° for the automatic and hand-scored methods respectively. Table 1 shows the mean saccade lengths to the cups by search condition for each of the scoring methods. The saccade lengths to the cups in the two search conditions were compared using a one-way analysis of variance (ANOVA). The mean saccade length to the cups was reliably greater in the matched cup condition than the mixed cup condition using the hand-scoring method, $F(1, 18) = 5.36$, $MSE = 25.45$, $p < .05$, and marginally greater using the automated method, $F(1, 18) = 3.85$, $MSE = 16.23$, $p < .10$. The mean fixation durations on the cups were 267 ms and 284 ms for the automated and hand-scored methods respectively. Table 1 also shows the mean fixation durations on cups by search condition. Although fixation durations were numerically greater in the mixed condition with both methods, these differences were not reliable ($Fs < 1$).

Table 1. Mean saccade lengths and fixation durations for cups derived by the automated and hand-scoring methods

	Condition	Automated	Hand-scored
Saccade length (deg)	Matched	8.8	9.0
	Mixed	12.3	14.2
Fixation Durations (msec)	Matched	258	272
	Mixed	275	294

3.2 Saccade Length Distribution

Figure 4 shows the overall distribution of saccade lengths generated by the automated scoring method for the entire viewing episode. The mean and mode of the distribution were 12.8 and 6.8 respectively. The mean saccade lengths did not differ between the two search conditions, $F(1, 1087) = 2.37$, $MSE = 107.77$, $p = .12$.

The automated and hand-scoring methods were cross-evaluated by comparing the set of saccade lengths that were measured using both methods. That is, the saccade lengths to the cups derived by the automated method were compared to the same saccades derived by the hand-scored method. Figure 5 shows the distributions of saccade lengths to cups derived from each of the two methods. As can be seen in the figure, the mode was higher using the automatic method (7.4°) than it was using the hand-scoring method (3.3°). The two distributions were compared using a repeated measures ANOVA. The overall mean based on hand-scoring (11.7°) was greater than the overall mean derived automatically (10.6°), $F(1, 88) = 6.38$, $MSE = 8.35$, $p < .05$. However, the difference between the means was quite small as was the size of the effect (partial eta squared was .068).

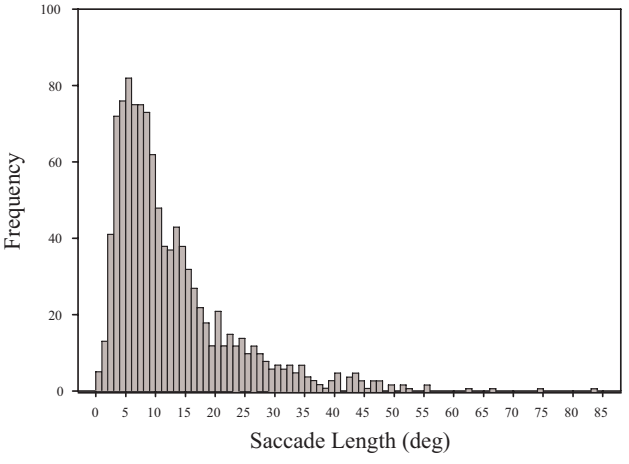

Fig. 4. Saccade length distribution for the entire viewing episode derived from the automatic scoring method

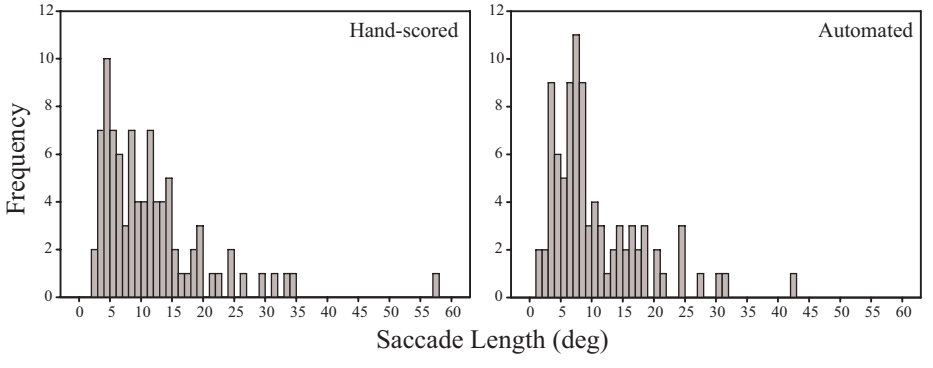

Fig. 5. Distributions of saccade lengths to the cups for the automatic and hand-scoring methods

Overall, the agreement between scoring methods was quite good. To begin, the saccade lengths derived from the two methods were reliably correlated ($r = .886$, $p < .001$). To determine quantitatively the nature of the relation between the two measures, the automatic scores were entered into a linear regression analysis with hand scores used as the predictor variable. The slope and intercept terms were both reliable, $b_1 = 0.76$, $p < .001$, and $b_0 = 1.75$, $p < .01$, respectively. A visual inspection of the distributions suggested that the largest difference between the distributions was the elevated number of saccades in the 7° range using the automated method. There were just as many short saccades in the automated distribution as there were in the hand-scored distribution. Thus, the mean was higher in the hand-scored distribution because of the increased number of longer saccades as opposed to a decreased number of shorter sac-

cades. Indeed, the regression suggests that the difference between the methods increases as saccades lengths increase with the automated method producing smaller estimates of saccades in the upper range of the distribution.

3.3 Fixation Duration Distribution

Figure 6 shows the overall distribution of fixation durations derived by the automated scoring method. The mean and mode of the distribution were 210 ms and 133 ms respectively. Fixation durations were greater overall in the mixed cup condition (mean = 220 ms) than in the match cup condition (mean = 199 ms), $F(1, 1109) = 10.31$, $MSE = 12334.59$, $p < .01$. As in the saccade length analysis, the automated and hand-scoring methods were cross-evaluated by comparing the fixation durations on the cups derived by the automated method to the same fixations derived by the hand-scored method. Figure 7 shows the distributions of fixation durations on cups derived by each of the two methods. As can be seen in the figure, the mode was 200 ms using both scoring methods. The two distributions were compared using a repeated measures ANOVA. The overall mean based on hand-scoring (281 ms) was marginally greater than the overall mean derived automatically (265 ms), $F(1, 88) = 2.70$, $MSE = 4053$, $p = .10$.

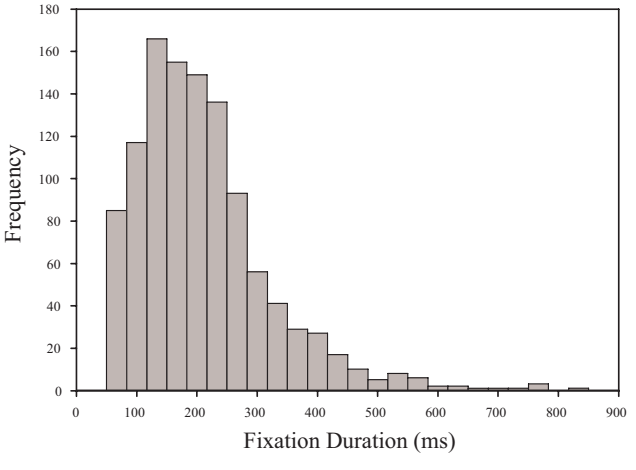

Fig. 6. Fixation duration distribution for the entire viewing episode derived from the automatic scoring method

Overall, the agreement between scoring methods was quite good. The fixation durations derived by the two methods were reliably correlated ($r = .748$, $p < .001$). Again, to determine quantitatively the nature of the relationship between the two measures, the automatic scores were entered into a linear regression analysis with hand scores used as the predictor variable. The intercept term

was not reliable when it was included in the model, $b_0 = 24.67$, $p = n.s.$, and the slope was 0.93 ($p < .001$) when the model was run without an intercept term. Visually, the largest difference between the distributions was the higher peak with the hand-scored method and the fact that there were twice as many fixations below 200 ms using the automated method. There were, however, no major systematic differences between the distributions.

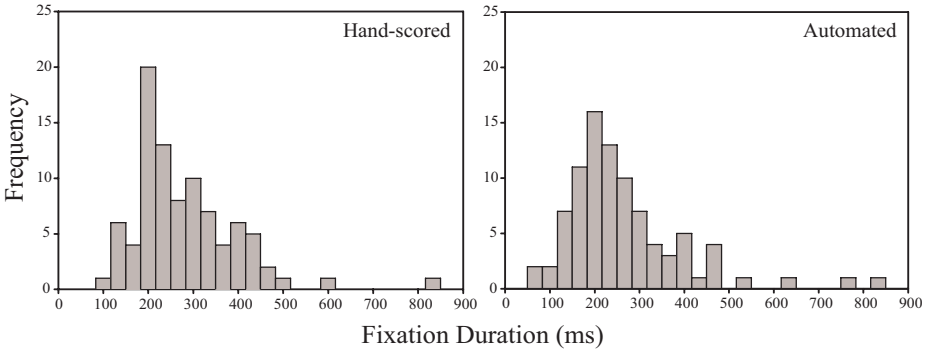

Fig. 7. Distributions of fixation durations on cups for the automatic and hand-scoring methods

4 Discussion

The main objective of this chapter was to begin to explore how eye movements are used to understand real-world environments. Up until now, the majority of research has been directed toward eye movements over scenes depicted in static displays. We believe that an important next step in scene perception research involves observing and understanding how the visual system behaves in real-world settings. To begin to examine this issue, we employed a visual search task, a task that is commonly used with static displays. Within the context of this task, we examined basic eye movement parameters to evaluate the generality of these measures in naturalistic settings.

The saccade length distributions generated by the present study provide a useful tool for evaluating real-world eye movement behavior. Our results yielded a greater number of long saccades than one might predict from earlier work. Both the saccade lengths to cups, as well as the distribution of saccade lengths overall were considerably larger than those reported in studies using static displays. The critical question is whether the differences in saccade length distributions are due to real differences in eye movement behavior versus due to differences in the equipment used to measure the behavior. A criticism of head-mounted eyetracking is that the spatial resolution does not allow for the detection of very small intra-object saccades. With the most precise stationary eyetrackers, however, saccades as small as 0.5° are very frequent. Therefore, the abundance

of longer saccades in our data could be driven by an insensitivity to smaller saccades.

There are a number of reasons to believe that this is not the case. First, the automated method allowed us to capture the eye movement behavior directly from the eye video rather than at a post processing stage. Since error would be likely to occur at each stage of processing, the spatial resolution issues that are present in the scene video are likely to be amplified relative to that provided by the eye video alone. In addition, the difference between automated and hand-scoring methods was greater at the high end of the distribution. The agreement between methods at the low end of the distribution suggests that even the scene video captures the lower end of the distribution reasonably well.

Second, if the mode of the distribution was elevated by an insensitivity to small saccades, we would expect the saccade length distribution to be truncated at the location of the lower limit of our equipment's spatial resolution. However, the pre-modal frequency of saccade lengths is greater than one would expect if small saccades are being lumped into fixations, and the post-modal tail of the distribution extends much further than it does in distributions derived from stationary eyetrackers on static displays. Visually, the similarity of the shape of this distribution to that found using static displays suggests that the distribution has actually shifted - reflecting an overall difference in viewing strategy between real-world settings and static displays. In addition to saccade lengths, we compared distributions of fixation durations. Fixation durations in the present study were appreciably smaller than those reported using stationary eyetrackers and static displays. However, as with the saccade length distribution, it is important to determine the extent to which differences might be attributed to equipment rather than behavior. One problem that arises in the generation of fixation distributions results from the abundance of head movements that occur during the viewing episode. Though the algorithm used by the automated scoring method was designed to robustly dismiss the slow eye movements that stabilize the POR in real-world coordinates while the head is moving, the potential for inappropriately splitting fixations remains. The fact that there were twice as many short saccades with the automatic method compared to the hand-scoring method supports this conclusion. This would not be an issue for the hand-scoring method because this is based on the stability of the POR in world-centered coordinates as reflected in the scene video. The automated scoring method, on the other hand, is based entirely on the stability or instability of the eye independent of this frame of reference.

A second issue for interpreting fixation durations comes from the limited temporal resolution of the HMET system. While the most resolute stationary eyetrackers are capable of sampling at rates greater than 1000 Hz, the video output of the HMET system is captured at a rate of 30 Hz (which corresponds to 33 ms frame samples). This difference would serve to artificially shorten our fixation durations because fixations that terminate just outside the sample window would be trimmed by as much as 33 ms. Given this amount of measurement error, we can only say with confidence that the mode lies between 133 ms and 166 ms and

the mean lies between 210 ms and 243 ms. While this mean is comparable to means found using stationary eyetrackers and static displays, the frequency of smaller fixations could reflect real differences in viewing behavior between our task and the tasks used in earlier studies. This could reflect the influence of time pressure, but further research is necessary to disentangle this result.

In addition to issues of the generality of eye-movement behavior during real-world viewing, our experiment was designed to determine whether differences in the specificity of the stimulus of interest modulates the perceptual span. The hypothesis was that by providing an example of the cup in advance the visual system would be better tuned to detect features in the environment that are likely to belong to the search target, thereby increasing the perceptual span. Under the assumption that saccade lengths reflect perceptual span, the saccades to the cups were expected to be greater in the matched cup condition. The saccade lengths to the cups observed with the automated and hand-scoring methods were both in agreement with this prediction. While this result lends support to the idea that the visual system is designed to adapt to specific task-related constraints, the fact that cups were the same color in the matched condition and of varied colors in the mixed condition prohibits a strong interpretation because the items were not matched for saliency. The fact that overall fixation durations were greater in the mixed condition is consistent with this possibility. Nevertheless, the reliable difference in saccade lengths between conditions demonstrates an ability to detect differences in perceptual span with the present methodology and suggests a promising avenue for future research.

5 Summary

The control of eye movements during scene perception is a topic that continues to gain interest in the scientific community. Our goal in this chapter was not only to begin to test the generality of eye movements in real-world environments but also to illustrate the advantages and challenges associated with head-mounted eyetracking. In the study of scene perception, what we really want to know is how the eyes behave during naturalistic viewing situations, where head movements are allowed and the field of view is unconstrained. Yet in practice we often use relatively constrained settings to make inferences about how people view the real world. One of the unique abilities that a head-mounted eyetracker affords over a stationary eyetracker is complete freedom of movement on the part of the participant. One could therefore argue that this methodology results in a more accurate portrayal of eye movement behavior in the real world. However, the complexity of the data analysis poses a considerable challenge. Eye movements are complex enough when the head is stationary and the view is constrained. Discerning meaningful patterns from this complex behavior is compounded by the dissociation of reference frames introduced by the freedom of head movement. As a result, the utility of HMET depends greatly on the advancement of analytical methods to deal with the added complexity.

The differences observed in the present study relative to those found using static displays suggests the enterprise of advancing these methods is warranted, and the degree of convergence between the automated and hand-scoring methods suggests the automated methodology is a promising start. There are a number of avenues that remain to be explored. For example, quantifying head movements has the potential to increase the accuracy with which saccades and fixations are determined because head movements seemed to occur within every observed video frame in the present study. However, because the abundance of head movement could be a product of the time pressure imposed in the current task, future methods should be tested in the context of alternative tasks. While utilizing HMET in a search task provides an initial glance, exploration with other tasks is necessary to provide a complete picture of eye movement control in real-world settings.

In conclusion, the data reported here suggest both similarities and differences between the viewing of natural environments and the viewing of static displays. Saccades lengths were longer than typically observed using static displays, but fixation durations appear to generalize across viewing situations. Finally, the scaling of saccades with the precision of the search target's representation provides yet another example of the visual system's ability to adapt and make efficient use of the information it is given.

Acknowledgments

This work was supported by the National Science Foundation (BCS-0094433) and the Army Research Office (W911NF-04-1-0078), and by a National Science Foundation IGERT graduate training grant (ECS-9874541). The opinions expressed in the article are those of the authors and do not necessarily represent the views of the department of the Army or other governmental organizations. Reference to or citations of trade or corporate names do not constitute explicit or implied endorsement of those entities or their products by the authors or the Department of the Army. Facilities for this research were supported by the Center for the Integrated Study of Vision and Language. We wish to thank Karl Bailey, Monica Castelhano, Dirk Colbry, and Nan Zhang for their contributions to the work presented.

References

1. Henderson, J. M.: Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), Eye movements and visual cognition: Scene perception and reading (pp. 260-283). New York Springer-Verlag (1992)
2. Henderson, J. M., Pollatsek, A., & Rayner, K.: Covert visual attention and extrafoveal information use during object identification. Perception & Psychophysics, 45, (1989) 196-208
3. Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. Perception & Psychophysics, 57, 787-795.

4. Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. Vision Research, 35, 1897-1916.
5. Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. The Quarterly Journal of Experimental Psychology, 38A, 475-491.
6. Antes, J.R. (1974). The time course of picture viewing. Journal of Experimental Psychology, 103, 62-70.
7. Buswell, G. T. (1935). How people look at pictures. Chicago: University of Chicago Press.
8. Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. Perception and Psychophysics, 2, 547-552.
9. Yarbus, A.L. (1967). Eye Movements and Vision. New York: Plenum Press.
10. Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. Eye Guidance in Reading and Scene Perception. Edited by: Underwood, G., Elsevier Science Ltd. 269-293.
11. Henderson, J. M. (2003). Human gaze control in real-world scene perception. Trends in Cognitive Sciences, 7, 498-504.
12. Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. Spatial Vision, 10, 165-188.
13. Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. Perception, 26, 1059-1072.
14. Parkhurst, D., & Niebur, E. (2003). Scene content selected by active vision. Spatial Vision, 16, 125-154.
15. De Græf, P., Christiæns, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. Psychological Research, 52, 317-329.
16. Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. Journal of Experimental Psychology: Human Perception and Performance, 25, 210-228.
17. Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. Journal of Experimental Psychology: Human Perception and Performance, 4, 565-572.
18. Castelhano, M. S., & Henderson, J. M. (2003). Flashing scenes and moving windows: An effect of initial scene gist on eye movements [Abstract]. Journal of Vision, 3(9), 67a, http://journalofvision.org/3/9/67/, doi:10.1167/3.9.67.
19. Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? Vision Research, 41, 3559-3565.
20. Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372-422.
21. Pelz, J. B., Canosa, R., Lipps, M., Babcock, J., & Rao, P. (2003). Saccadic targeting in the real world [Abstract]. Journal of Vision, 3(9), 310a, http://journalofvision.org/3/9/310/, doi:10.1167/3.9.310.