Portable Eyetracking: A Study of Natural Eye Movements

Jeff B. Pelz*, Roxanne L. Canosa, Diane Kucharczyk,
Jason Babcock, Amy Silver, and Daisei Konno

Visual Perception Laboratory
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, NY 14623

ABSTRACT

Visual perception, operating below conscious awareness, effortlessly provides the experience of a rich representation of the environment, continuous in space and time. Conscious visual perception is made possible by the 'foveal compromise,' the combination of the high-acuity fovea and a sophisticated suite of eye movements. Our illusory visual experience cannot be understood by introspection, but monitoring eye movements lets us probe the processes of visual perception. Four tasks representing a wide range of complexity were used to explore visual perception; *image quality judgments, map reading, model building,* and *hand-washing.* Very short fixation durations were observed in all tasks, some as short as 33 msec. While some tasks showed little variation in eye movement metrics, differences in eye movement patterns and high-level strategies were observed in the model building and hand-washing tasks. Performance in the hand-washing task revealed a new type of eye movement. 'Planful' eye movements were made to objects well in advance of a subject's interaction with the object. Often occurring in the middle of another task, they provide 'overlapping' temporal information about the environment providing a mechanism to produce our conscious visual experience.

Keywords: eye tracking, visual perception, complex tasks, natural tasks, portable eyetracking, wearable eyetracker, map reading, task-dependence, fixation durations, planful eye movements

1. BACKGROUND

Despite the seeming ease with which we perceive the world around us, visual perception is actually a complex process that occurs at a level below conscious awareness. Our conscious perception of the environment is that of a high-resolution, large field-of-view scene, continuous in space and time. However, because available neural resources cannot sustain such a representation, this perception is illusory. Like many 'optical illusions,' this is the result of the brain's attempt to make sense of the world with only partial information available.

Because the process of perception occurs below the conscious level, it does not yield to introspective report. However, monitoring observers' eye movements during a task can provide a tool to better understand visual perception. The retina in the human eye exhibits extreme anisotropy in photoreceptor density, with a small central region of high-resolution (the fovea) surrounded by a periphery with much lower spatial resolution. This design requires a mechanism for moving the eyes rapidly to direct the line-of-sight toward objects of interest several times per second. Limited spatial resolution in the periphery forces the oculomotor system to execute eye movements in tasks that require high spatial resolution. In natural environments eye movements are also made toward task-relevant targets even when high spatial resolution is not required. Such 'attentional' eye movements, made without conscious intervention, can reveal attentional mechanisms and provide a window into cognition.

*Correspondence: pelz@cis.rit.edu

In the past, researchers have sought to understand visual perception by breaking complex behaviors into simple, isolated *micro-tasks* then designing analogs of these task elements that can be studied more easily. The micro-tasks were then performed under unnatural (laboratory) conditions, typically with a small number of potential targets. This historical approach was justified with the argument that complex tasks could be understood as the sum of simpler sub-tasks. These experiments were performed because of the difficulty of monitoring complex tasks performed in their native environments.

While much was learned by the study of micro-tasks in the past, it is now clear that complex tasks are more than the sum of simplified sub-tasks. In fact, it is impossible to break down tasks with high-level cognitive components into meaningful elements without losing the very nature of the task under study [see, *e.g.*, Collewijn, *et. al.*, 1992, and Kowler, *et. al.*, 1992]. This argument about the weaknesses of examining micro-tasks under laboratory conditions is important, but the difficulty of understanding complex tasks at higher levels must be acknowledged. Because so much of what we accomplish in 'everyday' complex tasks is performed without conscious intervention, it is very difficult to describe via introspective report. This is especially true for over-learned tasks, such as driving to work, or performing a familiar task that is part of a daily routine. In both cases, subjects tend to describe only large-scale goals (*e.g.*, "check for approaching cars at the intersection," or "wash your hands"). Details about the strategies actually employed to accomplish the seemingly simple, over-learned tasks do not reach consciousness. If the method of conscious report is excluded because of its inability to capture important elements of complex tasks, we are forced to search for another tool.

Humans execute well over 100,000 rapid eye movements ('saccades') per day. The vast majority of these complex movements are programmed and executed without being conscious of the goal of each saccade [Becker, 1991]. If we are to understand vision, we need to understand the role that eye movements play. Yet until now eye movements have been treated largely as simple 'reflexive' movements made in reaction to stimuli in the environment, justifying experiments performed with a small number of 'point' targets flashed rapidly before an observer. Experiments conducted in this mode, which humans resort to only when stripped of virtually all visual stimuli, offer little insight into the performance of natural tasks. If we no longer treat eye movements as reflexive reactions to the environment, but rather recognize them as a window into cognition, we can use them to study the underlying attentional mechanisms and cognitive processes they reveal. While acknowledging that the task of extracting underlying strategies by observing behaviors is difficult [see *e.g.*, Viviani, 1990], there is increasing evidence that the approach yields important insights into behavior [Kowler, 1990, Ballard, *et al*, 1995].

The 'job' of the oculomotor system is two-fold; 1) to stabilize retinal images maintaining high spatial acuity in the face of observer and/or object motion, and 2) to target different objects and/or regions by moving the eye to ensure that the image of the new target is projected onto the fovea. It is this second class that we are concerned with here, so only brief descriptions of the image stabilization eye movements are given below.

The first class of eye movements maintains clear vision by stabilizing the retinal image. Such stabilization also assures that the image of an object or region in the center of the field-of-view is kept over the fovea. Sophisticated mechanisms exist to accomplish this goal in the face of eye, head, body, and object motion. These eye movements are often grouped into four categories [Carpenter, 1988]:

i) The *vestibular-ocular reflex (VOR)* rotates the eyes to compensate for head rotation and translation. Rotational and linear acceleration are detected by the semicircular canals and otolith organs in the inner ear. The resultant signals are used to command compensating eye movements.

ii) *Optokinesis* stabilizes the retinal image caused by large-field motion. Retinal slip induced by field motion is used to initiate eye movements at the appropriate rate to cancel out image motion.

iii) *Smooth-pursui*t eye movements are similar to optokinesis, but allow arbitrarily sized targets to be stabilized instead of large-field motion. A moving target is required for smooth eye movements; the eyes cannot move smoothly across a stationary object.

iv) *Vergence* eye movements counter-rotate the eyes to maintain the images of an object at a given depth to be maintained at corresponding locations on the two retinae.

The second class of eye movements performs the critical task of moving the eyes to a new object or region of interest, in essence destabilizing the retinal image. These *saccadic* eye movements are rapid, ballistic movements that move the eyes to new targets, typically identified in the periphery of a previous view. Saccades are made to objects requiring the high acuity afforded by foveal acuity or to 'attentional targets.' It is these eye movements that are of the most interest in our research because they provide an externally visible marker of the manner in which visual attention is deployed in the environment.

Much of the research on eye movements to date has been focused on understanding the mechanics and dynamics of the oculomotor system. The question of how successive fixations are aligned spatially has also received much attention. Most of this research has been aimed at discovering how the visual system 'knows' where the eyes are situated for each fixation so that the individual images captured with each fixation can be correctly aligned to build the rich internal representation we experience. Evidence is emerging, however, that we may have been asking the wrong question. We are able to use regularities in the environment to maintain a stable representation without resorting to complex alignment mechanisms [O'Regan, 1992, Pelz and Hayhoe, 1995], and large changes in the environment may go undetected [Simons and Levin, 1997]. Understanding visual perception requires us to ask a similar, but orthogonal question about the *temporal* stitching of successive views. This issue has not arisen with experimental tasks in the past because task complexity was purposely restricted.

We are studying eye movements in complex tasks and natural environments so that we can better understand the *process*, rather than the mechanics, of visual perception. An important goal of the research described here is to devise tasks that elicit natural eye movements. In order to investigate the manner in which vision is used in support of higher-order goals and tasks, we examined four tasks; *i) image quality judgments*, *ii) map reading*, *iii) model building*, and *iv) hand-washing*. The tasks were selected because they represent a wide range of task complexity.

In the least complex task, *image quality judgments*, subjects were instructed to rate a series of images at varying levels of JPEG compression. Subjects viewed the images in 'soft-copy' and 'hard-copy' forms, rating each on a subjective image quality scale. This task is considered the least complex because it requires only oculomotor control and verbal response. No other motor movements are required; the field-of-view was small enough that most gaze changes were completed without significant head movements.

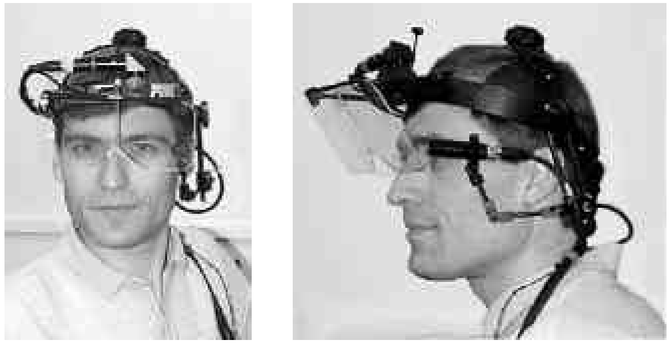
In the next task, *map reading*, subjects searched for two locations on a printed road map, then indicated the path between the two points. Map reading was considered more complex than the first task because map reading includes several sub-tasks not required in the image quality judgment task. Map reading required visual search over a wide field (with head and whole body movements), eye-hand coordination for indicating locations and tracing a path between locations, and the cognitive load in remembering locations and plotting the course constrained by the road map.

The third task, *model building*, required subjects to construct a model rocket following written and illustrated instructions. This task was more complex than the map reading task because it included visual search within an extended three-dimensional space, physical manipulation of pieces in the model kit, and following detailed, sometimes confusing directions.

In the final task, *hand-washing*, subjects entered a washroom to wash and dry their hands. While the task may at first seem simple, it requires a number of sophisticated, high level sub-tasks. The complexity of over-learned tasks, such as hand-washing or driving, is often apparent only during learning. After performing a task many times it becomes automated to the extent that the inherent complexity is no longer apparent. The hand-washing task required subjects to move under visual guidance; search for and manipulate objects, and perform a number of relatively complex hand, arm, trunk, and whole body movements.

2. METHODS

Subjects' eye movements were monitored in each task with a video-based infrared eyetracker based on the Applied Science Laboratories E5000 eyetracker. This device monitors eye position by tracking the pupil center and the first-surface corneal reflection of an infrared illuminator. A CCD camera aligned with the illuminator images the eye. Figure 1 shows the headgear used with the ASL eyetracker.



a) *b)*

Figure 1 ASL eyetracker headgear. The module above the headband in *a)* contains the IR illuminator and eye camera. The visor and scene camera are visible in *b)*.

The headband-mounted ASL eyetracker was not adequate for the hand-washing task, because the subject had to move freely into and within a room. A 'wearable eyetracker' was developed in the Visual Perception Laboratory at RIT to allow eyetracking in a broad range of natural tasks. As seen in Figure 2, the eyetracking system was built into goggles with the control system and power supply carried in a backpack.

a) *b)*

Figure 2 RIT Wearable eyetracker *a)* tracker hardware backpack *b)* custom headgear

Video records of eye movements were analyzed with a computer-controlled VCR. The videotape could be moved forward or backward at variable speed, and moved frame-by-frame under computer control. A frame-accurate timecode was automatically read by the lab computer so that fixation and eye movement events could be recorded along with the time at which they occurred. A data file containing timecode, event coding, and elapsed time from the start of the trial was stored. The angular extent of each saccade and the duration of the intervening fixations was determined for each trial. Because the subjects' head movements were not constrained by the eyetracking device, saccade size was scored as the shift in gaze (the sum of eye-in-head and head-in-space). Fixation durations were scored at video-frame resolution (33 msec) except for very short fixations. In the case that gaze changed direction without an intervening fixation visible at the 33 msec frame-resolution, the video records were analyzed at video-field resolution (17 msec). Fixation durations were only scored as 33 msec if they were framed by eye movements in adjoining video fields.

3. EXPERIMENT I - IMAGE QUALITY JUDGMENT

In the *image quality judgment* task, subjects rated a series of color images at varying compression levels. Two conditions were used; in the 'softcopy' condition, subjects viewed the images on a 21" 5500K Mag DJ920 monitor. The images were displayed as 5" x 7.5" at 65.5 dpi. In the 'hardcopy' condition, 300dpi dye-sublimation prints of the same size were viewed in a D5000 light booth. The field-of-view was small enough that most gaze changes were completed without significant head movements. The visual field was restricted to the region containing the image, but movements were not restricted in any way. Subjects rated each image on a subjective image quality scale.

Figure 3 is a sample image used in Experiment 1 at a quality level of '40' of a possible 100. Figure 4 shows the image segment indicated in the previous figure at quality levels '0' and '10.' Subjects made a total of 50 quality judgments; 5 images x 5 quality levels x 2 conditions (hard-copy and soft-copy). Image and quality were randomized within trials, blocked by condition. Figure 5 shows the normalized quality rankings for two subjects.

Figure 3 JPEG sample image; 'quality' = 40

a) *b)*

Figure 4 JPEG sample image; *a)* 'quality' = 0 *b)* 'quality' = 10

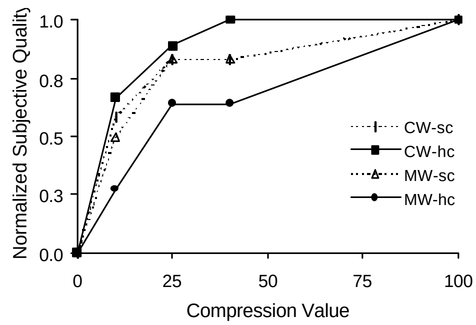


Figure 5 Normalized subjective quality rating *vs*. compression value for subjects CW and MW. Open symbols and broken lines are softcopy display ('sc'), solid symbols are for hardcopy ('hc').

Figure 6 shows the relative frequency of fixation durations for subject MW in the hardcopy and softcopy conditions. A shift toward shorter fixation durations in the hardcopy condition is evident. We considered the possibility that the difference was due to the order in which the conditions were blocked, but subject MW performed the hardcopy condition first, so if order was a significant variable one would expect shorter fixations on later trials after subjects become more familiar with the task and the images.

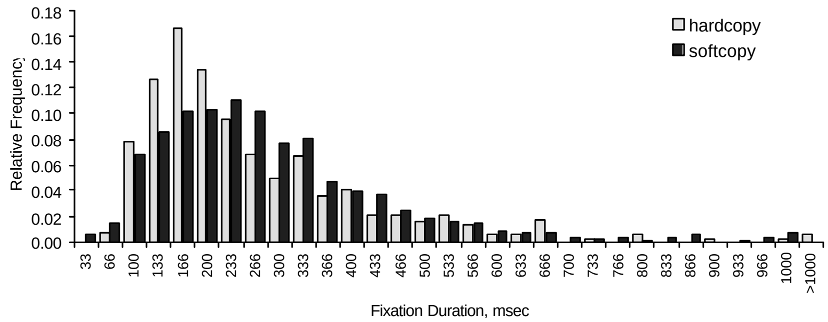


Figure 6 Frequency histogram of fixation durations for subject MW; hardcopy *vs*. softcopy

Unlike subject MW, subject CW did not show a significant difference between hardcopy and softcopy conditions. The relative frequency of CW's fixation durations were not statistically different than the distribution of subject MW's softcopy data. Figure 7 shows the relative distribution of fixation durations pooled across CW hardcopy, CW softcopy, and MW softcopy. While minimum fixation durations reported in the literature are typically $\geq$ 200 msec, 18% of the fixations in these trials were $\leq$ 166 msec.

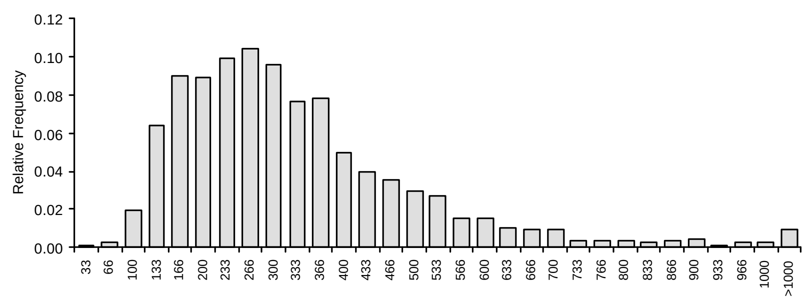


Figure 7 Frequency histogram of fixation durations pooled over subjects CW and MW softcopy.

Such short fixations have been reported in the past, but usually under one of two specific conditions. The first is the case where subjects fixate a target until a peripheral cue appears at one or more predetermined positions. The subject is instructed to move his/her eyes to the second point at target onset. In this case, socalled 'express saccades' are initiated with latencies of less than 200 msec. Express saccades can be considered 'predictive' saccades in the sense that the peripheral target appear only at known locations; attention can already be focused at those locations. The second case in which short fixations are reported is when the fixation occurs as part of a two-part saccadic sequence. In this case, the two saccades may be programmed at the same time, reducing the latency of the second saccade onset. While short fixations have been reported in natural tasks [Epelboim, *et al*, 1997, Hayhoe, 1999], both express and two-step eye movement patterns have been detected in laboratory conditions designed specifically to elicit short fixations. Neither case seems likely in the image quality judgment task; the subject is examining the image for signs of compression artifacts and/or other characteristics that influence image quality ratings. Presumably each fixation occurs so that task-relevant information can be extracted from the fixation points within the image.

While Figure 6 and Figure 7 illustrate the significant number of short fixations for both subjects in both conditions, the mean fixation duration (and number of saccades per second) is within the range typically reported. Figure 8 shows the mean fixation durations for subjects MW and CW in both conditions. Error bars indicate ± one standard error of the mean.

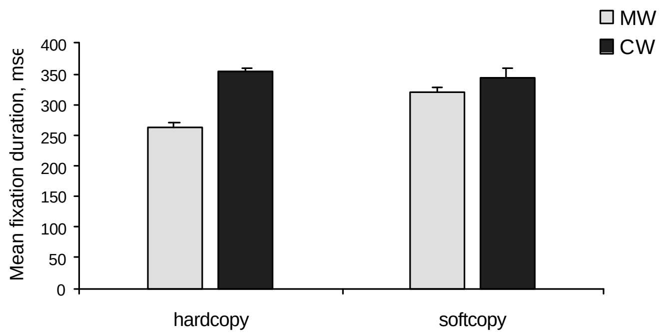


Figure 8 Mean fixation duration for hardcopy and softcopy conditions, subjects MW and CW

Figure 8 illustrates the significantly longer fixation durations subject MW had for the softcopy condition with respect to the hardcopy condition ($P < 0.05$), with ~20% increase at all quality levels. Subject CW did not show a significant difference between conditions.

Saccade size was also measured for the two conditions. The displays subtended approximately 20° at a viewing distance of 60 cm, though subjects were free to move and typically moved toward and away from the images in the course of a trial. Figure 9 shows the relative frequency of saccade size for softcopy and hardcopy conditions. The mean saccade size for the softcopy condition was significantly larger than the hardcopy condition (softcopy = 5.2° hardcopy = 4.3°, $P < 0.01$).

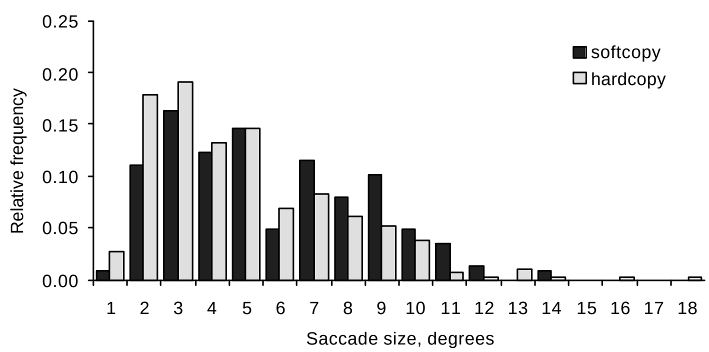


Figure 9 Relative frequency of saccade size for hardcopy and softcopy conditions

The results of Experiment I show that even for a simple task performed under natural conditions, the nature of eye movements can differ significantly from those seen under laboratory conditions.

4. EXPERIMENT II - SEARCHING A ROAD MAP

In the second task, *map reading*, subjects were instructed to search for two locations on a printed road map, then indicate the best path between the two points. Map reading was considered a more complex task than the first task because map reading includes several sub-tasks not required in the image quality judgment task. The map reading task required visual search over a wide field (many gaze changes required head movements), eye-hand coordination for indicating locations and tracing paths between locations, and the cognitive load associated with remembering locations and plotting the 'best route' connecting the two locations. The map was larger than the images used in the first task, and subjects typically used a shorter viewing distance when reading the map, so the map subtended a much larger visual angle; approximately 100° x 120°. The map reading task differed from the image quality task in another important respect. Rather than simply rank the image quality after free-viewing, subjects had to extract information from the map specific to each instruction, hold that information in memory, and plot the course between two points on the map.

The map reading task was performed in two segments. In the first segment (the 'search' sub-task), subjects were instructed by the experimenter to find two locations on a paper map (*e.g.*, "Charlotte," "Table Rock State Park"). They were told to use the map as they "normally would," and a magnifying glass was available for use if desired. After the subject found the two locations on the map, s/he was instructed to determine the best route between the two locations (the 'path finding' sub-task). Subjects traced the route on the map with their finger. Subjects displayed three distinct behaviors; conducting a *broad area* search over the entire map and a *narrow area* search over a region identified in the broad area search. The third behavior was tracing the best route between the two locations. The eye movement records were parsed into these behaviors (*broad search*, *narrow search*, and *path finding*) and analyzed as described above. Other behaviors (*e.g.*, referring to the map index, looking about the room, picking up the magnifying glass) represented a small fraction of each trial, and were excluded from analysis.

Figure 10 shows a highway map used in the task with a typical *broad area* search scanpath superimposed. The scanpath includes nine saccades and intervening fixations, representing approximately three seconds of search. Saccade size in this illustration ranges from 5° to 35°. Figure 11 illustrates typical scanpaths for a) *narrow area search* and b) *path finding* sequences. Eyetracking records for three subjects (authors JP and DK and naïve subject NS) participating in the *map reading* task were analyzed for fixation duration and saccade size.

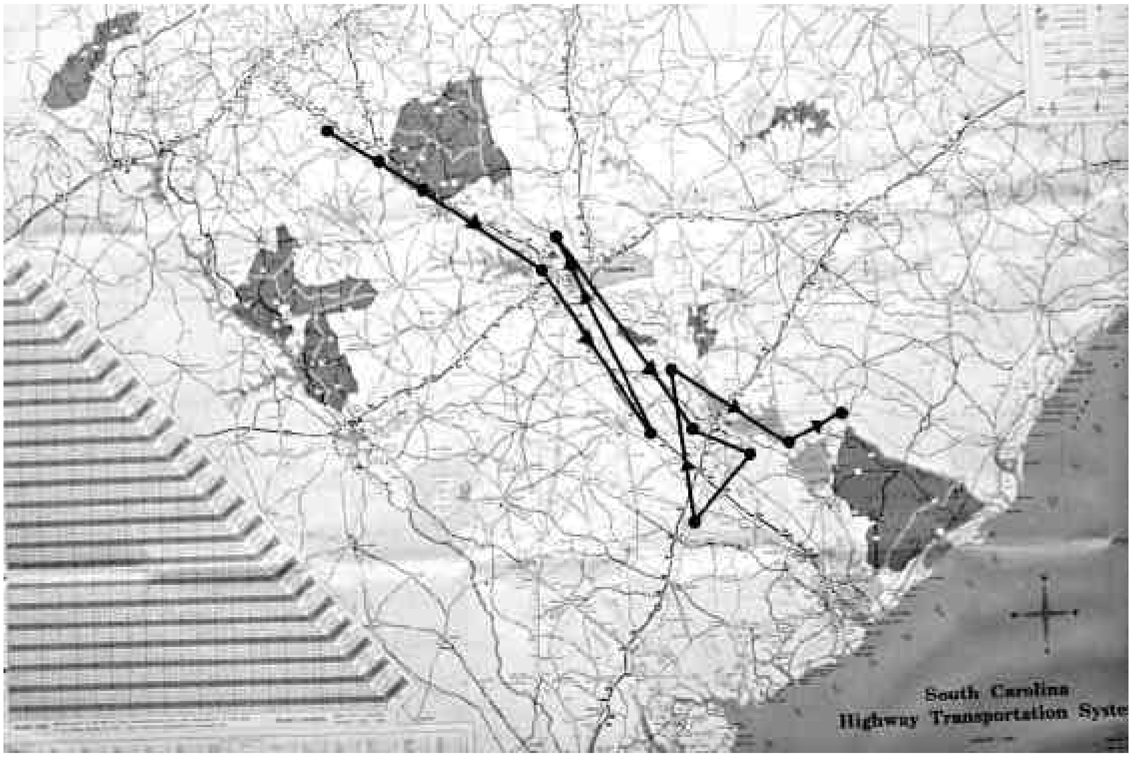
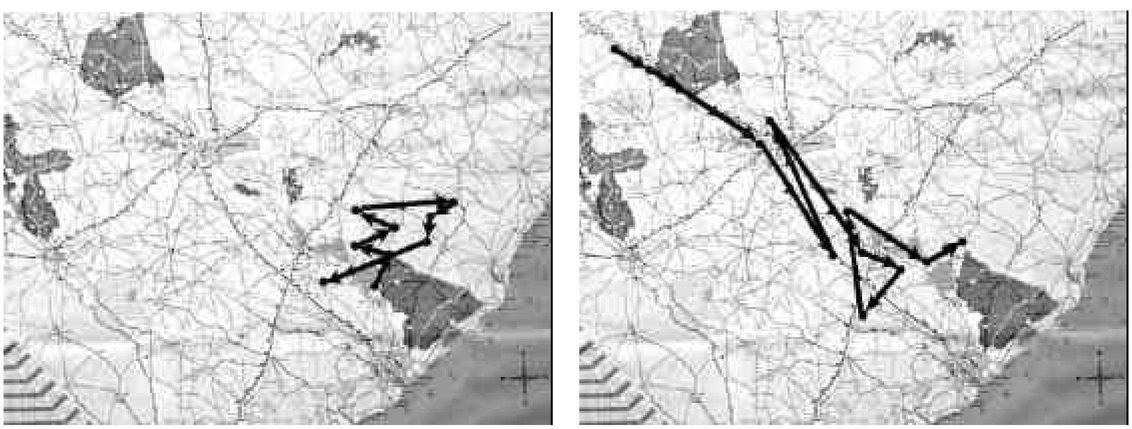


Figure 10 Typical *Broad search* scanpath in map reading task



a) *b)*

Figure 11 Typical *narrow search (a)* and *path finding (b)* scanpaths in map reading task

The frequency histogram of fixation duration during broad area search is shown in Figure 12. While there are some differences among the three histograms (notably the number of fixations longer than one second), the distributions of fixation durations are remarkably similar for all three sub-tasks. Figure 13 is the frequency histogram of fixation durations collapsed across the broad search, narrow search, and path finding conditions. As in the image quality judgment task, there were a significant number of very short fixations. The distribution mode was 166 msec, and almost half of the fixations were ≤ 200 msec. Even more surprising was that 12% of the fixations were ≤ 100 msec. While fixations this brief are not unusual in conditions designed to elicit 'express saccades,' they are shorter than those seen in most laboratory conditions. Recall that another condition under which very short fixations are seen is in the two-step fixation sequence.

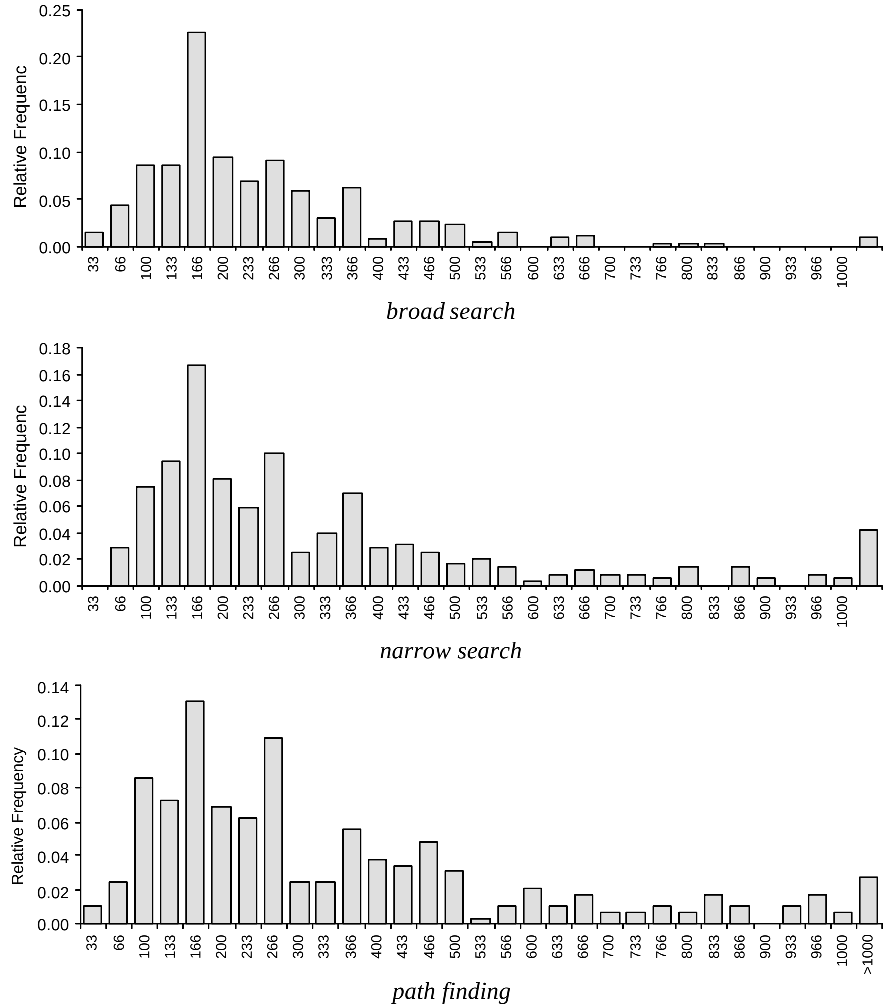


Figure 12 Relative frequency of fixation duration pooled for subjects DK, JP and NS for *broad search*, *narrow search*, and *path finding*

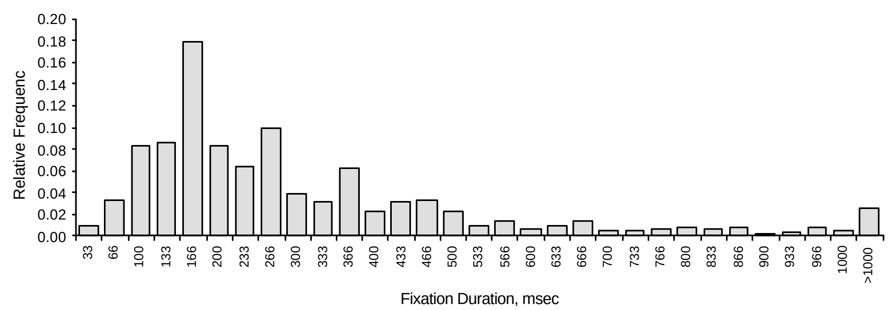


Figure 13 Relative frequency of fixation duration for subjects DK, JP, and NS collapsed across *broad search*, *narrow search*, and *path finding*

It would be surprising if such a two-step strategy were adopted in the broad or narrow search cases because the subject would be programming an eye movement without knowledge of what was found at the first fixation. An alternative hypothesis is that the short fixations are part of a 'corrective' sequence, occurring just before a small corrective saccade resulting from imperfect targeting of the initial saccade. This hypothesis was tested by examining the relationship between fixation duration and the size of the saccade following the fixation to determine whether the short fixations preceded short saccades. Figure 14 shows plots of saccade size *vs.* fixation duration for the *broad area* and *narrow area* searches, and *path finding*. While many of the shorter fixations were followed by small saccades in broad area, narrow area, and path finding, the correlation values were low (R^2 = 0.001, 0.004, 0.013, respectively), and in all cases there were many instances of short fixation durations followed by large saccades.

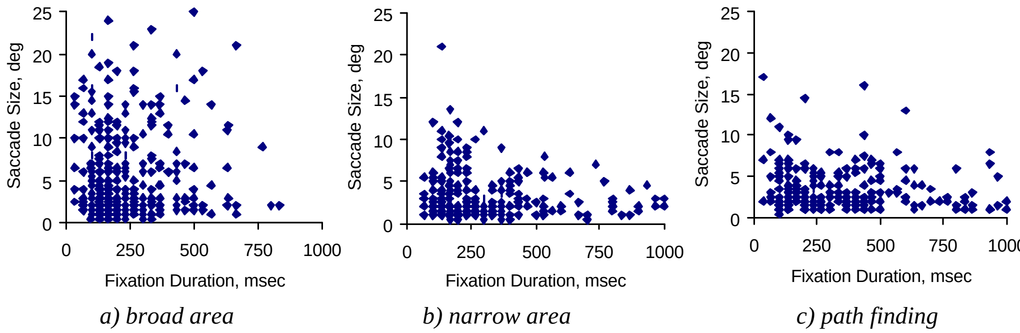


Figure 14 Saccade size *vs.* fixation duration for a) *broad area*, b) *narrow area* search and c) *path finding* data pooled for subjects DK, JP, and NS

Another indication that the short fixations are not just part of short multi-step sequences is that they occur in 'strings' of short fixations. Fixation strings of length two could represent a sequence of two saccades in which the second saccade is 'pre-programmed' concurrently with the first saccade. Fixation strings of length three and above are far less likely to be the result of pre-programming. Analysis of fixation records showed that many of the short saccades were indeed part of two-part strings, but there were also multiple occurrences of strings of three fixations ≤ 100 msec. If the threshold for a 'short' fixation is extended to 166 msec (still significantly shorter than the commonly reported 200 – 300 msec), there are several strings of short fixations of length four and five, and isolated strings of 6 - 8 fixations.

5. EXPERIMENT III - MODEL BUILDING

The map reading task was more complex than the image quality judgment task in several respects, but is still a relatively simple task when compared to tasks that are performed as part of everyday activities. The third task used in this research, *model building*, was selected to extend the complexity along several dimensions. The workspace was larger (a 1 m x 2 m tabletop, subtending nearly 180° horizontally), and three-dimensional objects had to be manipulated with fine motor control and eye/hand coordination. Subjects followed written instructions with illustrations. The kit consisted of many pieces in multiple packages that had to be sorted and searched to find the pieces needed for each segment of model building. Figure 15 shows a subject constructing a model rocket.

As in the first two tasks, subjects' eye movements were recorded as they performed the model building task. The task required several sub-tasks, many of which can be described at different levels. For the purpose of this experiment, we considered three sub-tasks; *reading, search,* and *manipulation* as illustrated in Figure 16. The black cross-hair in each image indicates the subject's gaze in the scene as she reads instructions, searches for a needed part, and constructs the model. The semitransparent image in the upper right corner is the image of the pupil from the eye camera. It is used to eliminate artifacts from blinks and to help find saccade onset and completion.

Figure 17 shows relative frequency histograms for two subjects in the reading, search, and manipulation sub-tasks, clearly showing the task-dependence of eye movement patterns in this task. Short fixations are not present in the manipulation task, evidently replaced by very long fixations. The search sub-task, on the other hand, shows a shift toward even shorter fixations, with the mode at 166 msec.

Figure 15 Model building task

a) "reading"

b) "search"

c) "manipulation"

Figure 16 Fixations in the building task were categorized as *reading (a)*, *search (b)*, or *manipulation (c).*

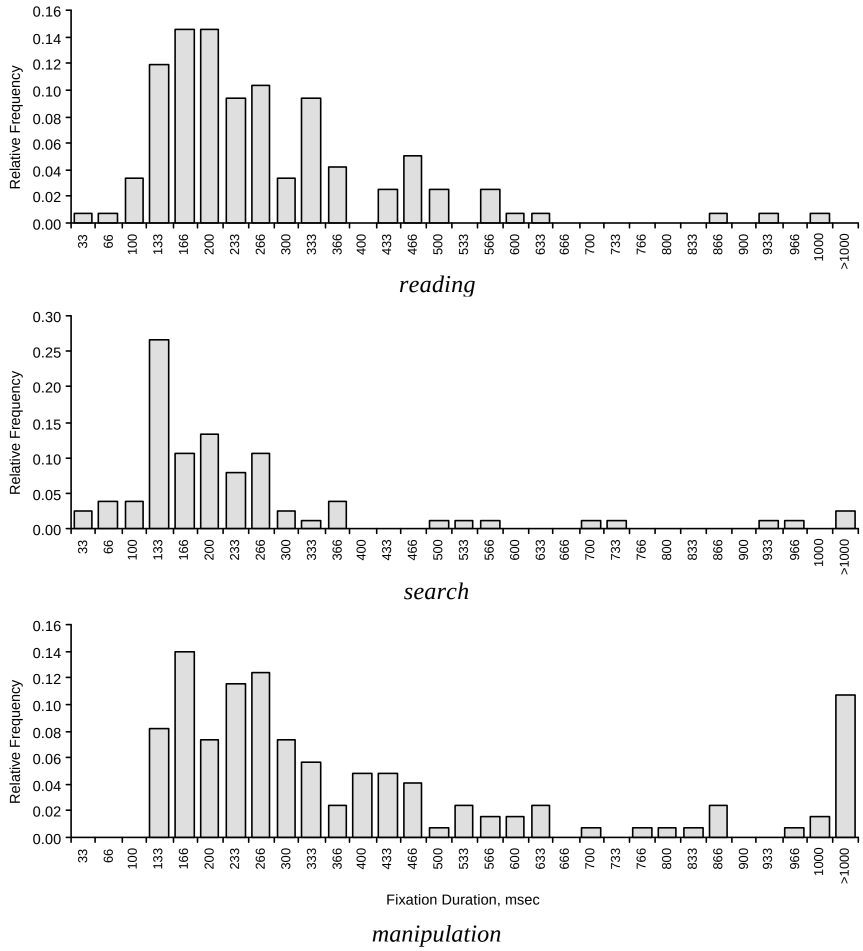


Figure 17 Relative frequency of fixation duration for subjects JB and JP for *reading*, search, and *manipulation*

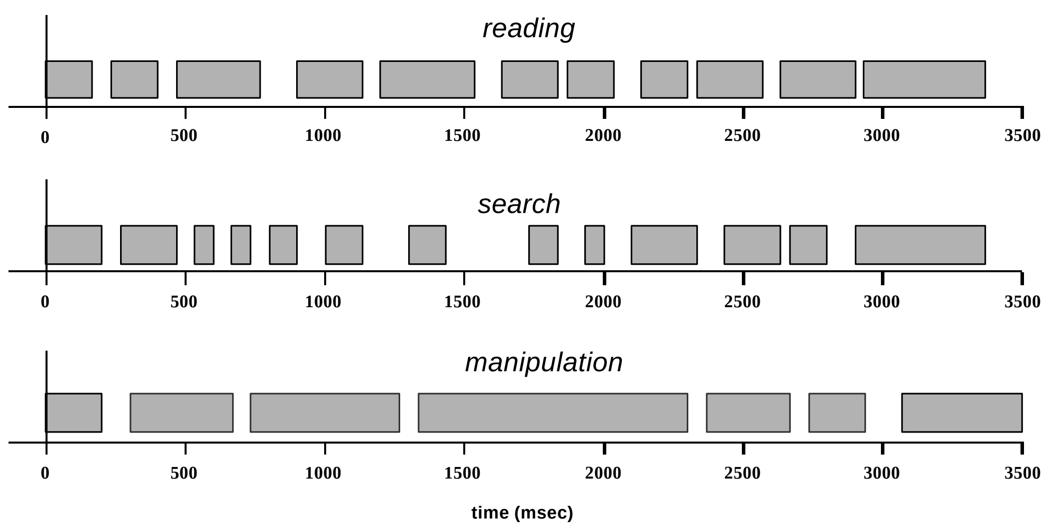


Figure 18 Fixation sequences for three sub-tasks in the rocket building task; *reading, search, and manipulation*. Bars indicate periods of fixation; spaces indicate gaze changes between fixation points.

There were also distinct differences in the temporal sequence of fixations. Figure 18 illustrates the differences in eye movement strategies employed in the three sub-tasks. The large number of very short fixations in the search sub-task are separated by large gaze shifts, often lasting longer than the fixations. In contrast, the reading pattern is more regular, with the familiar sequences seen in laboratory reading tasks. The manipulation pattern shows the long fixations common in this task seen in Figure 17.

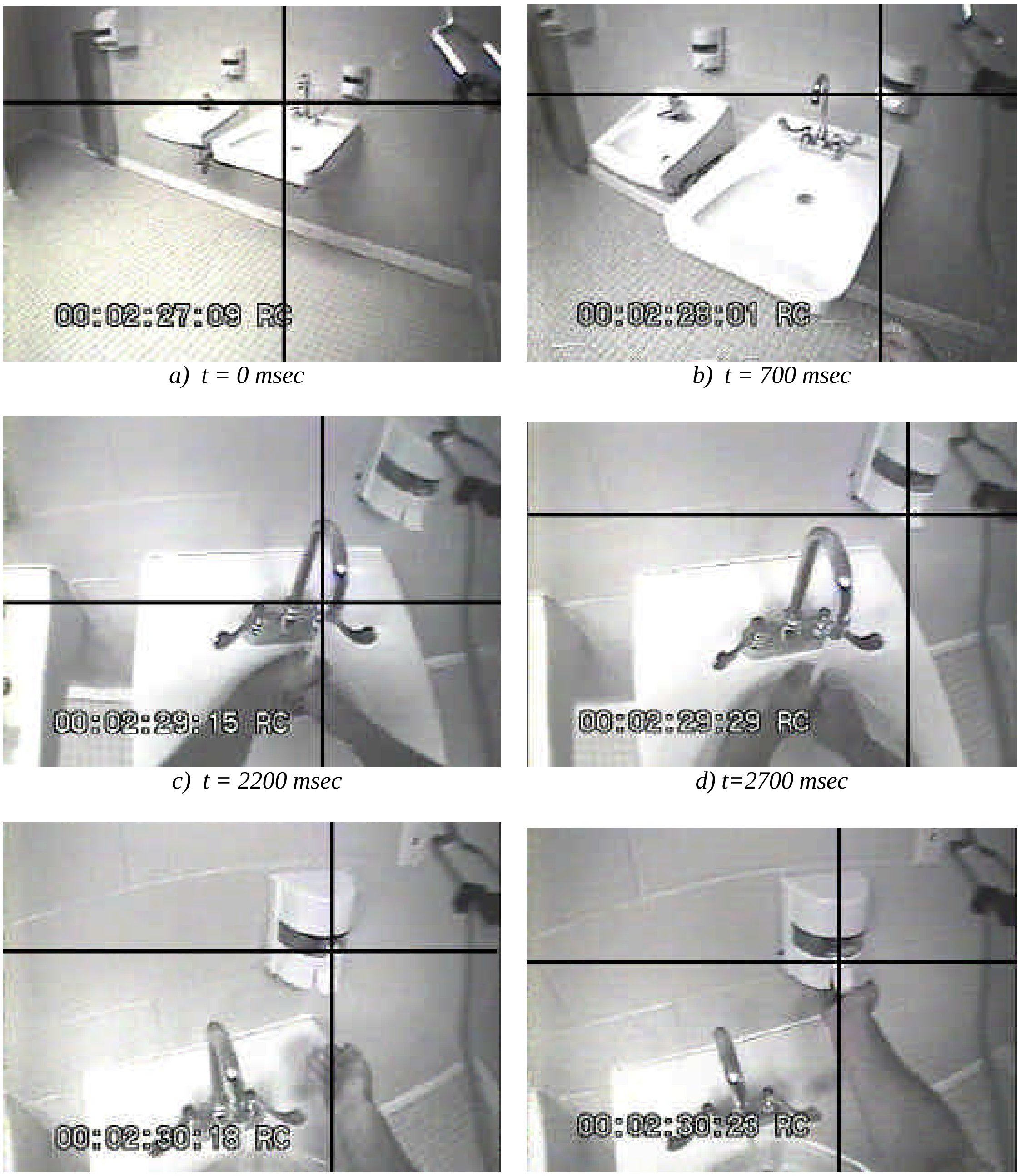
6. EXPERIMENT IV - HAND WASHING

In the *hand-washing* task, subjects were instructed to enter a washroom to wash and dry their hands. They were not given any other instructions, and were free to move at will. While the task at first seems simple, it requires a number of sophisticated, high level sub-tasks. The complexity of over-learned tasks (*e.g.*, hand-washing, driving) is often apparent only during learning. After performing a task many times it becomes automated to the extent that the inherent complexity is no longer apparent. The hand-washing task requires subjects to move under visual guidance; search for and manipulate objects (*e.g.*, water faucets, soap and towel dispensers, waste receptacles, door handles), and perform a number of complex eye, head, hand, arm, and whole body movements.

Subjects' eye movements were monitored from the time they entered the washroom until they left the room. Subjects moved over a distance of several meters, so it is difficult to specify a 'field-of-view' because the entire room was visible during the task. Figure 19 shows a sequence of frames from a trial. The crosshairs indicate the point of fixation in the scene, and the timecode shows hours:minutes:seconds:frames (one frame = 33.3 msec). The tapes were analyzed by recording each time an object was fixated, and when physical contact was made with that target. For example, Figure 19 shows a 3.5 second sequence in which the subject approaches the sink, turns on the water, reaches for, and contacts the soap dispenser.

Examination of Figure 19 reveals an interesting phenomenon seen in several instances in the hand-washing task. Figure 19 a) shows the initial fixation on the sink as the subject approaches. 700 msec later, before the subject has reached the sink, the subject fixates the soap dispenser above and to the right of the sink. Note that this fixation does not serve the immediate task (turning on the water faucets and wetting the hands), rather it is a 'preview' of information that will be needed in the future. In Figure 19 c), 1500 msec after the preview fixation, the subject is still fixating the sink. Figure 19 d) shows a typical 'targeting' fixation on the soap dispenser 2000 msec after the preview fixation, and 600 msec before the reach toward the soap dispenser. These targeting moves, occurring approximately 500 – 1000 msec before a reach, have been reported in other natural tasks [Epelboim, *et al*, 1997, Land *et al*, 1998, Land and Furneaux, 1997] and are typical of reaching tasks requiring visual guidance [Biguer, *et al*, 1982]. The initial preview fixation that occurred 2600 msec before the reach must serve another purpose altogether. We propose that these 'overlaps' in the sequence of fixations is evidence of a mechanism that provides conscious visual perception that is seamless in time as well as in space. While visual perception is in essence a sequential task, the preview fixations serve to stitch together information gathered at different times.

These 'planful' eye movements that lead to preview fixations well in advance of an action are different than the 'guiding' eye movements seen in previous tasks. They are also different than the 'predictive' eye movements seen in tasks requiring a subject to make eye movements to a known location. Planful eye movements aid perception in complex, multi-step tasks such as the hand-washing task. They have not been evident in earlier studies because typical experimental tasks were too simple.



a) t = 0 msec *b) t = 700 msec*

c) t = 2200 msec *d) t=2700 msec*

e) t = 3300 msec *f) t=3500 msec*

Figure 19 Fixation/action sequence in hand-washing trial;
a) initial fixation on sink, b) 'preview' fixation on soap dispenser, c) wetting hands,
d) 'guiding' fixation on soap dispenser, e) reaching toward soap dispenser, f) contact soap dispenser

Figure 20 shows preview and guiding fixations during a 27 second segment of a hand-washing trial. The length of the bars indicates the elapsed time between a fixation on an object and interaction with that object. The light bars indicate instances in which no preview fixations were made; the length of the bar simply represents the time the guiding fixation preceded action with that object. Groups of dark bars indicate instances in which there were fixations on an object before the targeting saccade immediately preceding the interaction.

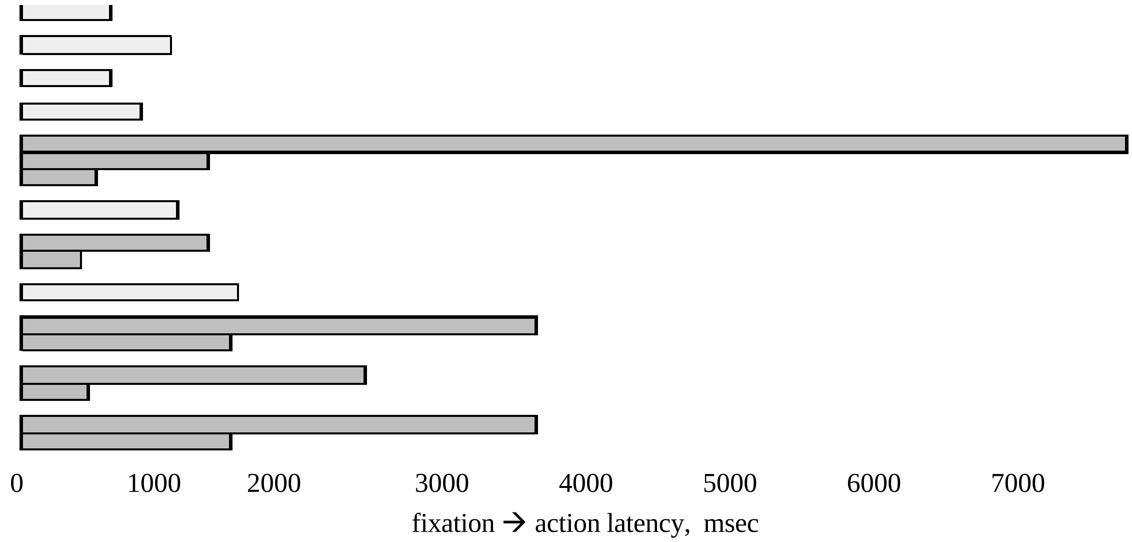


Figure 20 Fixation-to-action latency for 27 seconds of the hand-washing task
Length of bars indicate the elapsed time between an initial fixation on, and interaction with an object
Single bars (light shading) represent single fixation/action events
Contiguous bars (dark shading) represent multiple fixations on the same object

Although the figure shows the actions separately, the instances typically overlap in time. Figure 21 shows a sample timeline for a 20 second sequence. Objects in the scene are represented by horizontal sections; fixations on the object by narrow vertical extensions, and interactions by broad vertical extensions. Objects with only one fixation represent interactions with a targeting saccade preceding the reach. Objects with two fixations represent interactions in which a preview fixation was made in advance of the targeting saccade. Note the degree to which multiple objects are 'juggled' in the course of the task; two and sometimes three objects are fixated in a rapid sequence, indicating task-switching within the serial stream of fixations.

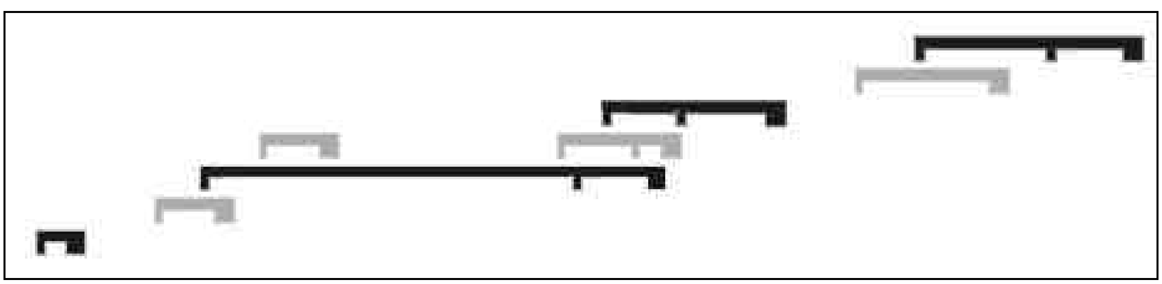


Figure 21 Fixation timeline for a 20 second sequence in the hand-washing task.

7. SUMMARY

Visual perception is a complex process that occurs below conscious awareness, making introspective report of limited use in attempts to understand the processes of perception. The success of the visual system in providing the illusion of a continuous spatial and temporal representation of the environment relies on a high-resolution fovea and an oculomotor system that alternately moves the fovea to sample new parts of the retinal image, and then stabilizes that image on the photoreceptor array. As a result, monitoring subjects eye movements as they perform natural tasks provides a way to probe the mechanisms of visual perception.

In the work reported here, four tasks were used to explore the role of eye movements in visual perception. The tasks; *image quality judgments*, *map reading*, *model building*, and *hand-washing* were selected because they represent a wide range of task complexity.

Surprisingly short fixation durations were observed in all tasks, with a significant number of fixations lasting ≤ 100 msec, with some as short as 33 msec. Fixation duration and saccade size varied little in the image quality task. Even in the map reading task, there was little variation among broad search, narrow search, and path finding. In contrast, when subjects performed the model building task, there were marked differences between reading, searching, and manipulation sub-tasks. The differences were seen in oculomotor metrics as well as in higher-level strategy shifts.

The most complex task, hand-washing, provided the most important result. A new type of eye movement was observed; 'planful' eye movements were executed to objects well in advance of interaction with the object. These eye movements often occurred in the middle of an ongoing task, providing 'overlapping' visual information about multiple targets. These eye movements may provide the mechanism that accounts for our conscious (and illusory) experience of a rich internal representation continuous in time and space.

Using the wearable eyetracker developed for this research to study humans performing complex tasks in natural environments opens up a new class of experiments that may help us better understand the processes of visual perception. Other experiments underway in RIT's Visual Perception Laboratory are examining the use of optic flow to guide locomotion in the face of eye and head movements, the relationship between task acquisition and eye movement patterns, and eye movements in many everyday activities, such as driving, casual conversation, and viewing works of art.

8. REFERENCES

Ballard, D.H., Hayhoe, M.M., and Pelz, J.B. (1995) Memory Representations in Natural Tasks, *Journal of Cognitive Neuroscience*, *7*, 68-82.

Becker, W., (1991) Saccades, in *Eye Movements, Vol. 8 in Vision and visual Dysfunction*, R.H.S. Carpenter Ed. CRC Press, Boca Raton.

Biguer, B., Jeannerod, M, & Prablanc, C. (1982) The coordination of eye, head, and arm movements during reaching at a single visual target. Experimental Brain Research, 46, 301-304.

Carpenter, R.H.S. (1988) *Movements of the Eye*. Pion Press, London.

Collewijn, H., Steinman, R.M., Erkelens, C.J., Pizlo, Z., & van der Steen, J. (1992) Effect of freeing the head on eye movement characteristics during three-dimensional shifts of gaze and tracking, Chapter 64 in *The Head-Neck Sensory Motor System*, Berthoz, A., Graf, W., & Vidal, P.P., Eds. Oxford University Press.

Epelboim J., Steinman R.M., Kowler E., Pizlo Z., Erkelens C.J., Collewijn H. (1997) Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Res. 37*: 2597-2607.

Hayhoe, M. (1999) personal communication.

Kowler, E. (1990) The role of visual and cognitive processes in the control of eye movements. In *Eye movements and their role in visual and cognitive processes*, Kowler, E. Ed. Elsevier Publishers

Kowler, E., Pizlo, Z., Zhu, G., Erkelens, C.J., Steinman, R.M., & Collewijn, H. (1992) Coordination of head and eyes during the performance of natural (and unnatural) visual tasks. Chapter 65 in *The Head-Neck Sensory Motor System, Berthoz*, A., Graf, W., & Vidal, P.P., Eds. Oxford University Press.

Land M.F., Furneaux, S. (1997) *The knowledge base of the oculomotor system*. Phil Trans R Soc Lond B 352: 1231-1239

Land, M.F., Mennie, N. and Rusted, J. (1998) Eye movements and the roles of vision in activities of daily living: making a cup of tea. Invest. Ophthalmol. Visual. Sci. 39: S457 ARVO Abstract 2094.

O'Regan, J.K. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology*, *46*, 461-488.

Pelz J.B. and M.M. Hayhoe, "The Role of Exocentric Reference Frames in the Perception of Visual Direction," *Vision Research*, **35**:2267-2275 (1995).

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science*, 1, 261-267.

Viviani, P. (1990) Eye movements in visual search: cognitive, perceptual, and motor control aspects. in Kowler, E. (ed.) *Eye movements and their role in visual and cognitive processes. Reviews of Oculomotor Research V4*, Elsevier, 353-383.