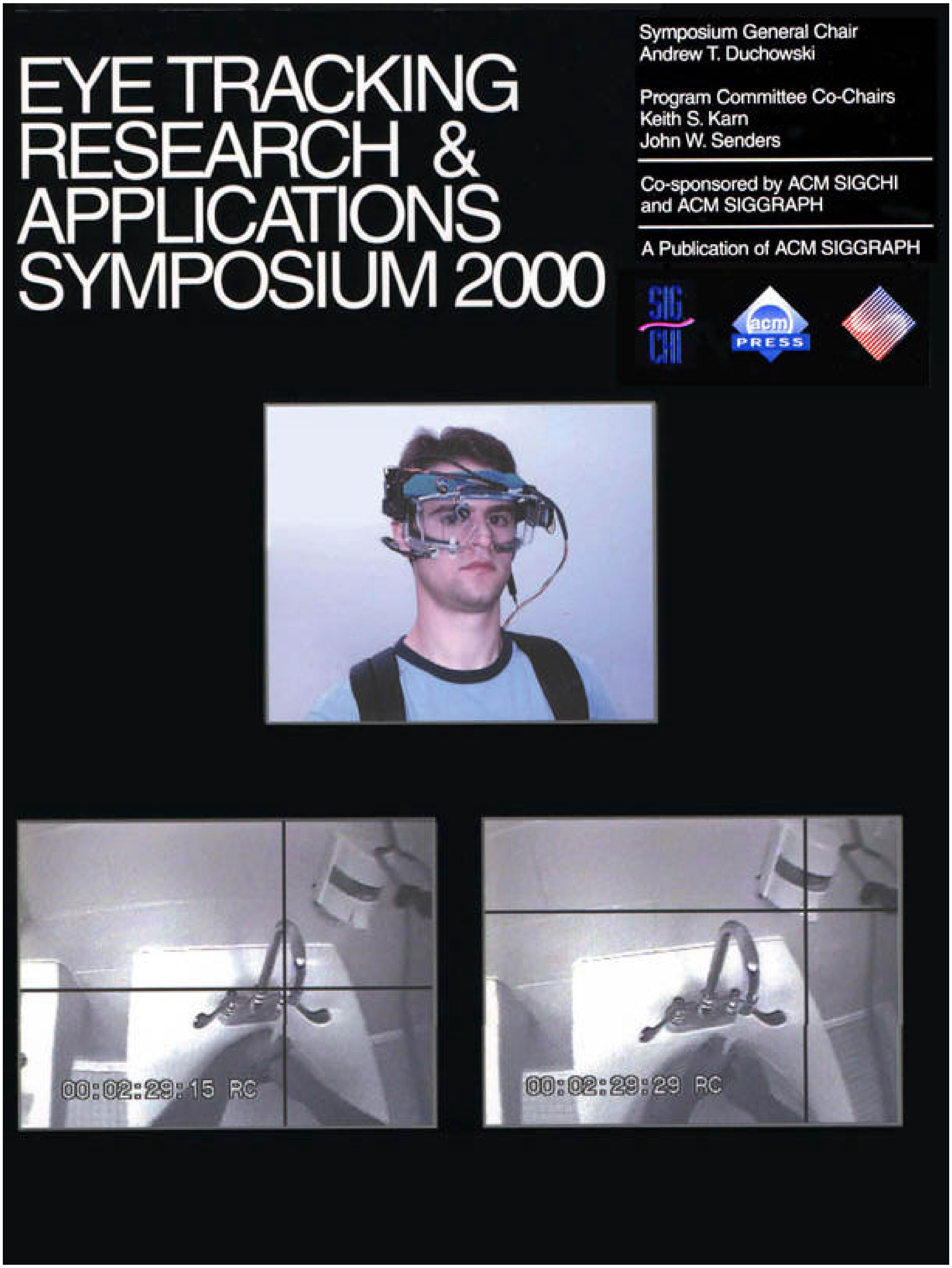
ACM SIGCHI Eye Tracking Research & Applications Symposium 2000



EYE TRACKING RESEARCH & APPLICATIONS SYMPOSIUM 2000

Symposium General Chair
Andrew T. Duchowski

Program Committee Co-Chairs
Keith S. Karn
John W. Senders

Co-sponsored by ACM SIGCHI
and ACM SIGGRAPH

A Publication of ACM SIGGRAPH

Extended Tasks Elicit Complex Eye Movement Patterns

Jeff B. Pelz[1], Roxanne Canosa, & Jason Babcock

Visual Perception Laboratory, Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, NY 14623 USA
[1] pelz@cis.rit.edu

ABSTRACT

Visual perception is an inherently complex task, yet the bulk of studies in the past were undertaken with subjects performing relatively simple tasks under reduced laboratory conditions. In the research reported here, we examined subjects' oculomotor performance as they performed two complex, extended tasks. In the first task, subjects built a model rocket from a kit. In the second task, a wearable eyetracker was used to monitor subjects as they walked to a restroom, washed their hands, and returned to the starting point. For the purposes of analysis, both tasks can be broken down into smaller sub-tasks that are performed in sequence.

Differences in eye movement patterns and high-level strategies were observed in the model building and hand-washing tasks. Fixation durations recorded in the model building tasks were significantly shorter than those reported in simpler tasks. Performance in the hand-washing task revealed *look-ahead* eye movements 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

Eyetracking, visual perception, complex tasks, extended tasks, portable/wearable eyetracking, fixation duration

1. INTRODUCTION

Despite the seeming ease with which we perceive the world around us, visual perception is a complex process that occurs at a level below conscious awareness. Subjective visual perception is that of a high-resolution, large field-of-view scene continuous in space and time. This percept, however, is actually an illusion. The anisotropic retina coupled with a sophisticated oculomotor system supports the apparently effortless perception of a continuous environment.

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. In natural environments, eye movements are 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.

The relatively simple tasks often used to investigate eye movements typically restricted subjects to a small number of targets, two-dimensional search, reading, image segmentation, *etc*. With few exceptions, experiments were performed by stationary observers viewing static scenes. In the majority of such tasks, the subject did not interact with the surrounding environment other than *via* verbal response or button-press. While such experiments have contributed much to our understanding of eye movement mechanisms, control structures, and metrics, they have provided little insight into human behaviors in the real world. 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.*, [3], and [6]). 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. 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.

The vast majority of the thousands of eye movements made daily are programmed and executed without consciously selecting the goal of each saccade [2]. Monitoring these eye movements has been a valuable tool in efforts to better understand visual perception, the associated attentional mechanisms, and cognitive processes. While acknowledging that the task of extracting underlying strategies by observing behaviors is difficult [12], there is increasing evidence that the approach yields important insights into behavior [5],[1].

The job of the oculomotor system is two-fold. One class of eye movements serves to stabilize retinal images to maintain high spatial acuity in the face of observer and/or object motion. 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 orient the eyes toward 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 to attentional targets 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 so that the 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 [9],[10], and large changes in the environment may go undetected [11]. 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 study the manner in which vision is used in support of higher-order goals and tasks. Two tasks were studied; the first task, *model building*, required subjects to construct a model rocket from a kit following written and illustrated instructions. The task required following detailed, sometimes confusing directions, searching a large workspace for small parts, and physical manipulation of pieces in the model kit.

In the second task, *hand-washing*, subjects entered a restroom to wash and dry their hands. While the task seems simple, it requires the execution of 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 control unit. This device monitors eye position by tracking the pupil center and the first-surface corneal reflection of an infrared illuminator. A video camera aligned with the illuminator images the eye. Figure 1 shows the headgear provided with the ASL eyetracker. The headband-mounted ASL eyetracker was not adequate for mobile observers. A wearable eyetracker was developed at RIT to allow eyetracking in a broad range of natural tasks. As seen in Figure 2, the wearable eyetracker uses lightweight CMOS video cameras to image the eye and scene. The eye and scene cameras are affixed to racquetball goggles, which are in turn connected to the battery-powered control unit. The control unit, the batteries, and two VTRs (to record the eye and scene video) were carried in a backpack worn by the subject. The eye and scene video records were merged, then the records 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 with a timestamp. Fixation durations were scored at video-frame resolution (33 msec) except for very short fixations, where the video records were analyzed at video-field resolution (17 msec).

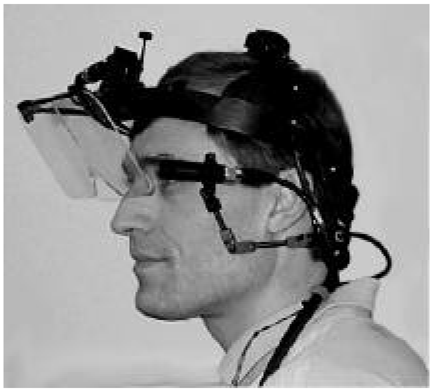


Figure 1 Head-mounted eyetracker used in Experiment 1

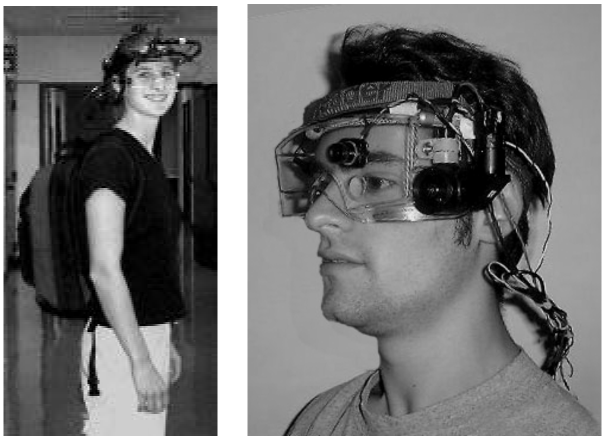


Figure 2 RIT Wearable Eyetracker used in Experiment 2

3. EXPERIMENTS

Two complex, extended tasks performed by freely moving observers were used in the research described here. In the first task, subjects built a model rocket from a kit following written and illustrated instructions. In the second task, subjects walked to a restroom, washed their hands, and returned to the starting point. In both tasks, subjects performed an extended task comprised of several lower level sub-tasks. Both tasks required significant interaction with the environment; in the model-building task seated subjects searched for and manipulated components needed to construct the model; in the hand-washing task, freely moving subjects moved through the environment, using stationary objects to complete the task.

3.1 Experiment 1: Model Building

In the first experiment, two subjects were fitted with the head-mounted eyetracker and seated in front of a 1m x 2m tabletop. A boxed model rocket kit, a pair of scissors, and a hobby knife were placed on the tabletop. The subjects were simply instructed to "make the model rocket," and progressed at their own pace. 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 3 shows a subject constructing a model rocket.

Figure 3 Workspace for Experiment 1; Model building

The model-building task can be considered as consisting of three sub-tasks. In the first sub-task, labeled *reading*, (see Figure 4) subjects read the instructions provided with the model rocket kit. The second sub-task, labeled *searching*, (see Figure 5) consisted of visual search of the workspace for pieces needed to complete the next step in constructing the model. In the third sub-task, labeled *manipulation*, (see Figure 6) subjects picked up the pieces located during the search phase and put them together with previously completed sections of the model.

The black cross-hair in each figure indicates the subject's gaze in the scene as she reads instructions, searches for a needed part, and constructs the model.

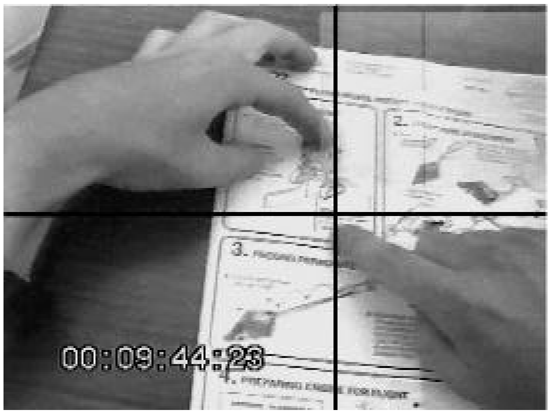


Figure 4 *Reading* sub-task in model-building task

Figure 5 *Searching* sub-task in model-building task

Figure 6 *Manipulation* sub-task in model-building task

The nature of the task dictated that the *reading, searching, and manipulation* sub-tasks be executed in repeating sequences. While at intermediate timescales the sub-tasks are serial, efficient progress could only be made by interleaving execution of the three sub-tasks for successive elements of the macro task.

The *reading* sub-task is similar to the reading paradigms used in other eye movement studies. The only distinguishing characteristic is that the subject was free to move the written instructions.

The *searching* sub-task is unlike traditional visual search experiments in which subjects are instructed to locate an isolated region or object with a specified characteristic on a two-dimensional display. Subjects had to search a large workspace for an object described or illustrated in the instructions. Once found, the piece was integrated into the model, removing it from the workspace.

The *manipulation* sub-task included picking up pieces and/or the partially completed model from the workspace, and constructing the model from those pieces. The instructions were ordered, but subjects were free to complete the task in any order they desired.

3.2 Experiment 2: Hand-washing

While Experiment 1 required subjects to perform an extended, complex task, it was not representative of typical everyday behaviors. In Experiment 2 subjects performed the overlearned task of hand-washing. Four subjects were instructed to walk to a restroom, wash their hands, and return to the starting point. Subjects' eye movements were monitored with the wearable eyetracker described in the METHODS section above.

While this is apparently a simple task, it requires a number of sophisticated, high-level sub-tasks. The complexity of such 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.

The subjects knew the location of the restroom relative to the starting point and were familiar with the layout of the restroom, so visual search in the traditional sense was not required. The hand-washing task could be broken down into twelve sub-tasks; *i)* walking to the restroom, *ii)* opening the door, *iii)* walking to the sink, *iv)* turning on the water, *v)* soaping the hands, *vi)* rinsing the hands, *vii)* turning off the water, *viii)* drying the hands, *ix)* discarding the towel, *x)* walking to the door, *xi)* opening the door, and *xii)* walking back to the starting point. These definitions are subjective and described at a relatively high level; each could be broken down further into lower level sub-tasks. The sub-task described above as "soaping the hands," for example could be broken down into a series of lower level tasks, such as a) locate the soap dispenser, b) reach for the soap dispenser, c) operate the dispenser lever, d) lather the hands, e) return the hands to the sink, *etc.* The sub-task level used were selected to support the analyses described in RESULTS below,

Subjects' eye movements were monitored from the time they began walking to the restroom until they returned to the starting point. Subjects moved over a distance of several meters, so it is difficult to specify a field-of-view because a large region in the building was visible during the task.

4. RESULTS

The videotaped records for each task were analyzed to determine fixation durations, gaze change size, and scanpath order. Gaze fixations were defined as any period in which a subject's gaze remained stationary with respect to an object in the field. Because the subjects were free to make unrestricted head and body movements, the eyes were frequently moving with respect to the head even during fixations. VOR and smooth pursuit eye movements that stabilized the retinal image often take place during the periods defined as fixations.

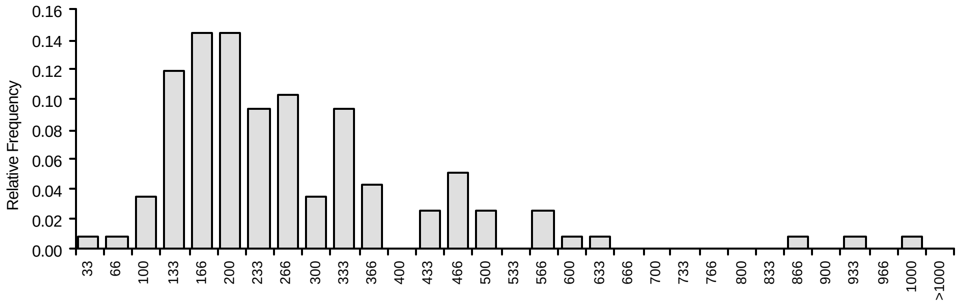
4.1 Experiment 1

There were distinct differences in the temporal sequence of fixations for the sub-task categories; *reading*, *search*, and *manipulation*. Figure 7 shows pooled relative frequency histograms of fixation duration in the model building task. Figure 8 illustrates typical fixation sequences for the three sub-tasks in the model building experiment; *reading*, *search*, *and manipulation*. The shaded bars indicate periods of fixation; spaces indicate gaze changes between fixation points. Short fixations are not present in the manipulation task, evidently replaced by very long fixations. The median fixation duration for the manipulation sub-task was 266 msec; the mean duration was 450 msec due to the relatively large number of very long fixations. The search sub-task, on the other hand, shows a shift toward shorter fixations, with the median at 166 msec, and mean of 275 msec. Note that the sequence of fixations is very different than that typically recorded in simpler tasks. In the *search* segments of Experiment 1, the durations of the gaze changes were often longer than the intervening fixations, though the extended gaze change at ~1500 msec in the search sequence shown in Figure 8 occurred during a large head movement, and is likely made up of a brief fixation bounded by two large gaze changes. Because no fixation meeting the two video-field criterion was found, the period appears as a single gaze change in the figure. The median value for fixation durations recorded during the reading sub-task was 200 msec; the mean was 275 msec.

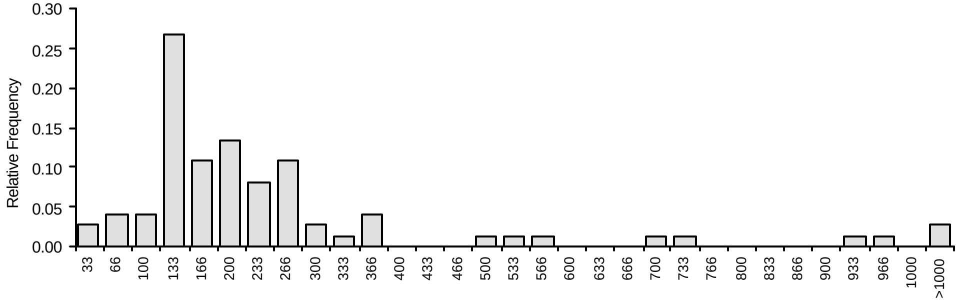
4.2 Experiment 2

The videotapes were analyzed by manually noting the timecode at which an object was fixated, and when physical contact was made with that target. Like the first experiment, Experiment 2 required subjects to perform a series of sub-tasks. Perhaps the most interesting result of Experiment 2 was the degree to which the mid-level sub-tasks were interleaved. Figure 9 illustrates the phenomenon; Figure 9a) shows the initial fixation on the faucets as a subject approaches the sink. Some 700 msec later, before reaching the sink, the subject fixates the soap dispenser above and to the right of the sink (see Figure 9b). Note that this fixation does not serve the immediate task (turning on the water faucets), rather it is a 'look ahead' to information that will be needed in the future. In Figure 9c), 1500 msec after the *look-ahead* fixation to the soap dispenser, the subject is still fixating the water faucets. Figure 9d) shows a typical guiding fixation on the soap dispenser 600 msec before the reach toward the soap dispenser, and 2000 msec after the look-ahead fixation. Subjects often made eye movements to the soap dispenser and towel dispenser while walking toward the sink, before the initial reach to the water faucets. These eye movements occurred several seconds before the reach toward the corresponding targets, and did not replace the guiding eye movements made ~500 msec before those reaches. The targeting eye movements, occurring approximately 500 – 1000 msec before a reach, have been reported in other natural tasks [4], [8], [7] and are typical of reaching tasks requiring visual guidance.

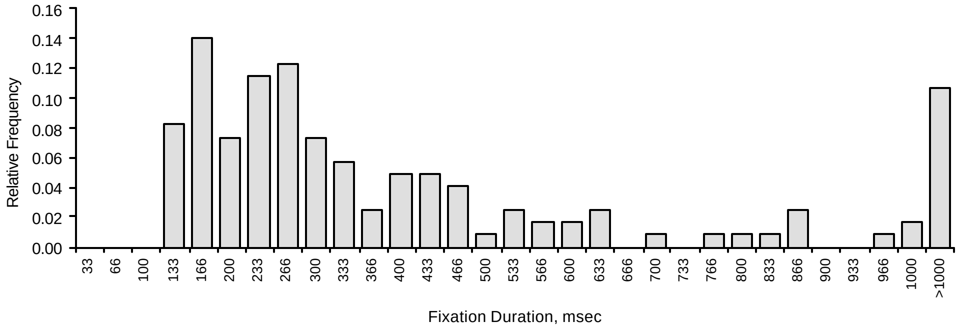
The initial look-ahead fixation that occurred 2600 msec before the reach must serve another purpose altogether. We propose that these overlaps in the sequence of fixations are evidence of a mechanism that provides conscious visual perception that is seamless in time as well as in space. The look-ahead fixations are not part of a conscious strategy; subjects were not aware that they were making the look-ahead eye movements. Analogous to the spatial extent of the retinal periphery, these look-ahead eye movements may support an orthogonal *temporal peripheral vision* that allows visual planning to extend into the future. All four subjects executed these look-ahead fixations at least once during the hand-washing task. On average, subjects made over three look-ahead fixations during the task, comprising approximately 3% of the total number of fixations recorded.



reading



search



manipulation

Figure 7 Fixation duration frequency histogram for *reading, search, and manipulation* segments

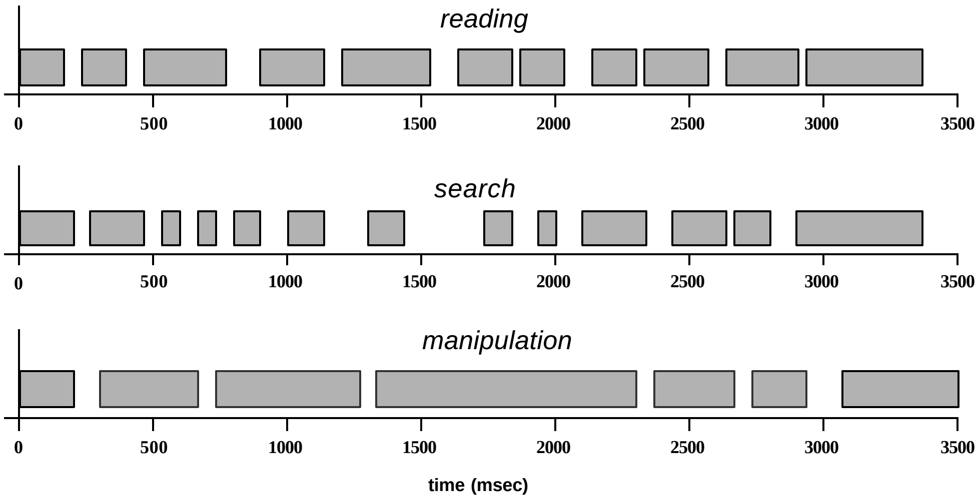
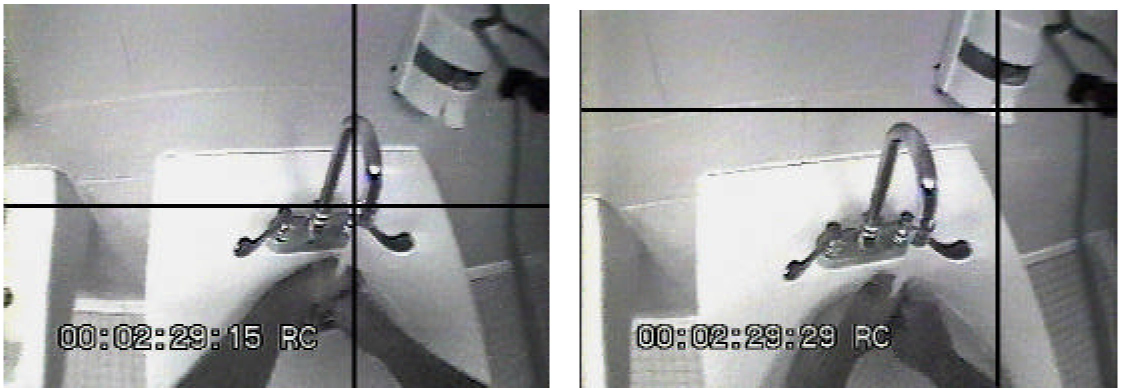
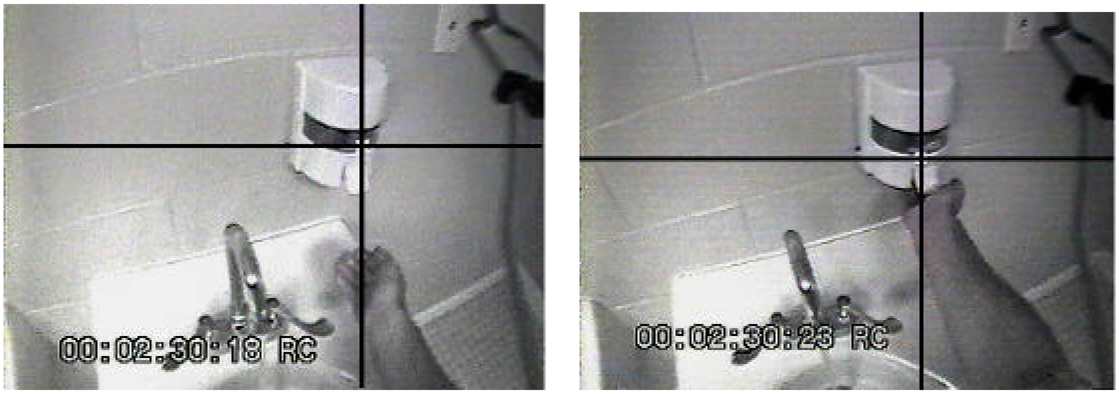


Figure 8 Fixation sequences for; *reading, search, and manipulation* segments (see text for details)

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 9 Fixation/action sequence in hand-washing trial;
a) initial fixation on sink, b) *look-ahead* fixation on soap dispenser, c) wetting hands,
d) guiding fixation on soap dispenser, e) reaching toward soap dispenser, f) contact soap dispenser

5. DISCUSSION

The results from both experiments demonstrate how profoundly eye movements are affected by an observer's task. The model-building task in Experiment 1 demonstrated the dramatic differences in eye movement patterns even between the three sub-tasks performed in the course of building the model. Fixation durations in the model-building task were markedly different from those reported in typical tasks with static scenes and/or stationary observers. Fixation durations less than 100 msec were not uncommon; the median fixation duration for the search task in Experiment 2 was 166 msec.

The hand-washing task revealed further complexities in oculomotor performance in extended tasks like those that make up daily life. The look-ahead eye movements observed in Experiment 2 were executed to objects well in advance of interaction with the object. These eye movements 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 (though illusory) experience of a rich internal representation continuous in time and space.

Using the RIT 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.

6. REFERENCES

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

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

[3] 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.

[4] 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.

[5] 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

[6] 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.

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

[8] 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.

[9] 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.

[10] 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).

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

[12] 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.

.