10-7-2001

Visual perception in familiar, complex tasks

Jeff Pelz

Roxanne Canosa

Jason Babcock

Jenn Barber

VISUAL PERCEPTION IN FAMILIAR, COMPLEX TASKS

Jeff B. Pelz†*, Roxanne Canosa, Jason Babcock, & Jenn Barber*

Visual Perception Laboratory, Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, NY 14623 USA †pelz@cis.rit.edu

ABSTRACT

Visual perception is virtually effortless, operating at a level below conscious experience. Because it most frequently operates without attentive intervention, it does not yield to introspective report. The foveal/peripheral distribution of photoreceptor density in the human eye requires a mechanism to rapidly re-target areas in the environment for visual inspection. The eyes are moved both toward areas where high-acuity, central vision is required and toward objects of interest to the current task. Monitoring those eye movements can provide a window into perception. Subjects' eye movements were monitored while they performed the familiar, complex task of hand-washing. Analysis revealed a novel perceptual strategy, in which objects of future interaction were foveated seconds before the information was "needed" for a task. These *look-ahead* fixations are task-dependent; while they occurred in over 3% of fixations in the hand-washing task, their frequency fell to less than 1% in a control experiment. We propose that the look-ahead fixations represent a strategic deployment of attentional and visual resources to optimize information gathering during natural tasks.

1. INTRODUCTION

Visual perception is virtually effortless for humans, yet attempts to design artificial systems that perform even rudimentary visual tasks have proven difficult. Perception unfolds over time and is at once a parallel and serial process. While the subjective experience of visual perception is that of a wide field of view, stable, high-acuity internal representation, mounting evidence suggests that it is far more limited and fleeting than it seems. The foveal/peripheral design of the human retina provides high acuity over a region subtending only about 0.1% of the full field. Acuity in the remainder of the field falls rapidly, supporting motion detection and acuity sufficient to identify regions for further investigation.

A sophisticated suite of eye movements is available to move the eyes rapidly to sample regions in the environment in rapid succession [2,6]. Because the eyes are directed towards targets of interest even if they do not require the high-acuity vision available in the center of the field, eye movement records provide a valuable tool to monitor and understand high-level visual perception [1,3]. For the most part, previous eye movement studies were limited to simple, single-step tasks. 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. The focus of the research described here is to study the manner in which vision is used in support of higher-order goals and tasks.

In colloquial use, the terms *complex* and *difficult* are synonymous and *simple* and *easy* are used interchangeably. In our work we make a distinction between the *simple-complex* and *easy-difficult* continua; indeed, they can be considered as orthogonal axes in task space. Even very complex tasks can be 'easy' given practice. Driving, for example, is difficult for the novice driver but is a familiar, over-learned task for experienced drivers. An experienced driver can drive a large distance over a familiar route without expending significant overt attentional resources. Perhaps a less obvious example is hand-washing. The ease with which we accomplish the task belies the inherent complexity of navigating a hallway, entering a washroom door, locating and manipulating water faucets, soap dispensers, hand-towels, and waste bin. The complexity would become apparent if one tried to program an autonomous robot to complete the same series of sub-tasks. In addition to the manual components of the task, attempting to navigate and perform the tasks based on an image stream from a mobile camera is beyond the state-of-the-art, despite the fact that a young child regards the same task as easy. Despite significant progress in traditional and active vision paradigms, artificial vision systems are still largely restricted to simple, repetitive tasks that they are explicitly programmed to complete.

0-7803-6725-1/01/$10.00 ©2001 IEEE

In the experiments described here, we used a custom-built, wearable eyetracker to monitor the eye movements of subjects as they performed the familiar, complex task of hand-washing. Observations regarding their behaviors revealed perceptual strategies that may support our illusory visual experience that is continuous in both space and time. In order to better understand the task-dependence of the perceptual strategies employed in complex tasks, a control task with the same low-level actions was included.

2. METHODS

2.1. The RIT Wearable Eyetracker

Subjects' eye movements were monitored using the *RIT Wearable Eyetracker*, a video-based, infrared eyetracker based on an ASL E5000 controller. The eyetracker monitors eye position by tracking the pupil center and the first-surface reflection from the cornea's surface. Figure 1 a) shows the custom headgear. Miniature CMOS video eye and scene cameras are affixed to racquetball goggles, which are in turn connected to the control unit and two VTRs (to record the eye and scene video) in a backpack worn by the subject, as seen in Figure 1b. The system computes eye position 60 times per second, and creates a video record of the scene from the subject's viewpoint, with the point of fixation indicated by crosshairs.

Figure 1 a) custom headgear b) backpack controller

2.2. Experiment 1: Hand-washing

Subjects performed the over-learned task of hand-washing. Subjects were instructed to walk to a washroom, wash their hands, and return to the starting point. Subjects' eye movements were monitored from the time they began walking to the washroom until they returned to the starting point.

2.3. Control experiment: Cup-filling

The control *fill-cup* task was identical to the hand-washing task except for the instructed goal. Subjects were given a cup and instructed go to the washroom, fill the cup with water, and return to the lab.

Up to the initial contact with the water faucet, subjects had to walk the same path, enter the same environment, and perform the same actions. At that point, the tasks diverged, so differences between the hand-washing and fill-cup tasks could be ascribed to a high-level strategy.

2.4. Data Analysis

The videotaped records for each task were analyzed to determine fixation durations, saccade amplitude, 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 occur during the periods defined here as fixations.

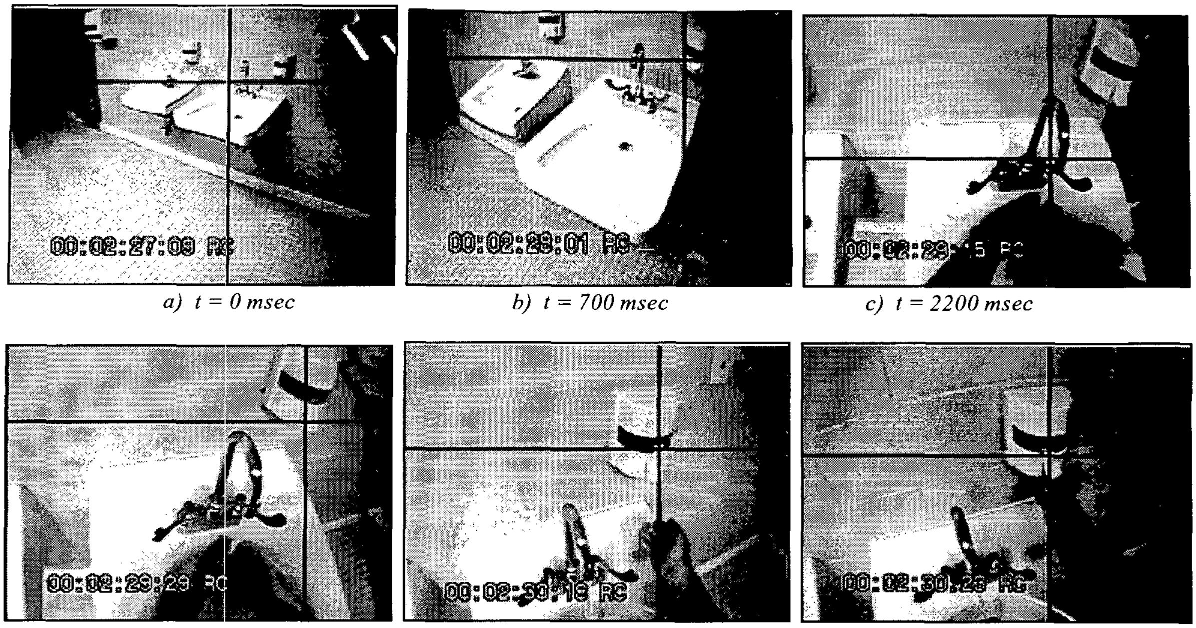
2.5. Subjects

Five paid undergraduate students participated in the experiment. All had normal or corrected-to-normal vision.

3. RESULTS

3.1. Experiment 1: Hand-washing

The task required subjects to perform a series of sub-tasks. An interesting result was the degree to which the mid-level sub-tasks were interleaved. Figure 2 illustrates the phenomenon; Figure 2a) 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 2b). 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 2c), 1500 msec after the look-ahead fixation to the soap dispenser, the subject is still fixating the water faucets. Figure 2d) shows a typical guiding fixation on the soap dispenser 600 msec before the reach toward the soap dispenser, and 2000 msec after the first 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. All subjects made look-ahead fixations at least once, and the look-ahead fixations made up over 3% of the total number of fixations recorded during the trial.



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 2 Sequence of video frames showing fixation points (indicated by crosshairs) during the task

3.2. Control experiment: Cup-filling

The low-level metrics of oculomotor behavior (*e.g.*, fixation duration and saccade amplitude) did not vary significantly between the main and control experiments. High-level behaviors, on the other hand, varied dramatically. Despite the fact that the hand-washing and cup-filling experiments shared nearly all the same component subtasks, the stated high-level goals of the tasks were different. The average number of fixations to the soap dispenser fell from over three in the hand-washing task, to less than one in the cup-filling task.

4. DISCUSSION

While visual perception is subjectively a parallel operation, vision is inherently serial. Because of the foveal/peripheral design of the human retina, the eye must be explicitly targeted to regions in the environment that requires high-acuity vision. Even when the peripheral view has sufficient acuity for a specific subtask, observers foveate regions of interest. The eye movements supporting the targeting serve to serialize complex tasks, and provide an externally visible marker of spatial and temporal attentional distribution.

With the observation that the look-ahead eye movements are common in some tasks, it is useful to consider a number of hypotheses regarding their utility in a given task. The eye movement patterns may be a general behavior elicited by low-level cues such as the conspicuity or salience of objects in the environment. We reject this hypothesis based on the results of the control condition, as the identical visual field was present in both tasks, yet the look-ahead fixations were nearly absent in the interval where the tasks were identical.

Alternatively, the look-ahead fixations may be part of a visual search process. Search may have two meanings in this context. The eye movements could be part of a general 'sweep' of the surround to identify objects in the environment, or they may indicate a targeted search for a specific object. The first alternative is rejected based on the observed pattern of fixations; while subjects do look at some objects as they enter the washroom, they do not initiate a sweeping sequence of eye movements, fixating on a large number of objects. Further evidence is provided by the results of the control condition, as the pattern is tied to the high-level goal rather than the environment. The second alternative is not supported by the eye movement patterns; subjects' look-ahead fixations to targets of future relevance are targeted accurately, and are not typically followed by corrective saccades centering on the object.

We propose that the look-ahead fixations provide a mechanism to 'stitch together' the stream of visual input resulting from the sequence of actions that make up daily life. Analogous to purported mechanisms supporting visual stability, look-ahead fixations may be part of a strategy that supports our subjective experience of an environment that is continuous in time as well as space. In the spatial domain, there is a large difference between the rapid sequence of retinal images due to eye movements and the subjective experience of a stable environment. The scene doesn't appear to jitter or jump as the retinal images certainly do, nor is there uncertainty about where the observer or scene is located in space. This phenomenon has drawn the attention of investigators for many years, and has led to several hypotheses about the mechanism(s) by which we achieve this visual stability. Knowledge of eye position *via* proprioception, efference copy of oculomotor commands, and regularities in the scene have all been shown to play a role in the perceived stability of the visual scene [5,6].

In the real world, we have another dimension to contend with; not only must vision provide spatial stability, it must also support temporal seamlessness. As we move through and interact with the world, building a stable representation is not sufficient. The visual system must supply a steady, reliable stream of information to support our conscious experience of an environment continuous in time and space. This issue has not arisen with experimental tasks in the past because task complexity was purposely restricted.

A look-ahead strategy could also ease the task-switching between the serial sub-tasks used to gather information from, and interact with, the environment. Like Land, Mennie, and Rusted's [4] report, we found that ~95% of fixations are dedicated to completing the immediate sub-task. The remaining 5% may be used to help bridge the task-switching that would otherwise be evident to subjects, as they are for complex tasks before we become proficient. Part of the transition from 'hard' to 'easy' in completing complex tasks is the gradual reduction in the conscious effort required to complete the discrete sub-tasks in sequence.

The strategy we are reporting only emerged when we observed subjects performing complex natural tasks in natural environments. Further exploration in domains where vision is examined in its native role as a tool supporting high-level perception is likely to identify other perceptual strategies that are fundamental to better understanding visual perception.

The hand-washing task revealed complex strategies in oculomotor performance in extended tasks like those that make up daily life. The look-ahead eye movements 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 demonstrate how profoundly task-dependent eye movements are, and may provide a mechanism that supports our conscious (but illusory) experience of a rich internal representation continuous in time and space.

5. 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] 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

[4] Land, M.F., Mennie, N., & Rusted, J. (1999) The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28: 1311-1328.

[5] 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).

[6] Steinman, R.M., Kowler, E., and Collewijn, H. (1990) New directions for oculomotor research. *Vision Research*, 30, 1845-1964.