10-3-2000

Eye movements and natural tasks in an extended environment

Roxanne Canosa

Follow this and additional works at: http://scholarworks.rit.edu/theses

Recommended Citation

This Thesis is brought to you for free and open access by the Thesis/Dissertation Collections at RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.
Eye Movements and Natural Tasks in an Extended Environment

Roxanne Canosa

B.S. State University of New York College at Brockport (1998)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Center for Imaging Science in the College of Science Rochester Institute of Technology

October 3, 2000

Signature of Author

Roxanne Canosa

Accepted by Coordinator, M.S. Degree Program

Date

Oct. 23, 2000
Eye Movements and Natural Tasks in an Extended Environment

I, Roxanne Canosa, hereby grant permission to the Wallace Memorial Library of RIT to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature of Author ____________________________________________
Roxanne Canosa

Date 11/2/00
M.S. DEGREE THESIS

The M.S. Degree of Roxanne Canosa
has been examined and approved by the thesis committee
as satisfactory for the thesis requirement for the
Master of Science degree in Imaging Science

Dr. Jeff B. Pelz, Thesis Advisor

Dr. Eriko Miyahara

Dr. Roger Gaborski

11/2/00
Date
Abstract

Eye movements can be thought of as a window onto pre-conscious thought. Patterns of visual fixations over time as well as space can reveal cognitive strategies that are not amenable to conscious control or verbalization. A spatial analysis of an eye movement trace usually emphasizes the role that eye movements have in moving the retinal image of an object of interest from the periphery to the fovea for closer inspection. It is generally believed that a sequence of fixations across a region of space builds up the perception of a high-resolution field of view everywhere. Recent studies have shown that this perception is largely illusory. The visual-perceptual system prefers to maintain a limited internal representation of physical objects in the world and uses the environment as an external source of information, accessing the information only at the time it is needed.

The goal of this research effort was to investigate the role that eye movements have in the performance of everyday tasks in a natural environment. A series of four experiments were conducted that represent an attempt to step away from the classical psychophysical approach of studying eye movements within the confines and control of the laboratory. There exists little precedence for this kind of approach, partly because past research efforts have emphasized a linear systems method to render the analysis tractable, and partly because the technology that is required to perform these experiments has not existed until recently. The hardware that was developed by the Visual Perception Laboratory at RIT specifically addresses the portability concerns that are crucial for successfully studying eye movements during natural tasks in a non-linear extended environment.

A model was developed to describe the temporal sequencing of eye movements in terms of a hierarchical structure of goal-oriented tasks, with individual fixations considered the lowest level of the hierarchy. The analysis gives evidence for the sequencing of eye movements based on a desire to maximize the efficiency of task performance over time by anticipating future activities. The purpose of this sequencing is to enhance interaction with the world under conditions of limited memory representations rather than to create the perception of a high-resolution field of view.
Acknowledgments

I would like to thank and acknowledge all those who helped to make this research project possible. Thanks to Jason Babcock for the countless hours spent developing and fine-tuning the eye-tracker, and for his help in recruiting and running subjects. Thanks also to Amy Silver for her coding expertise and her insights into what constitutes a good track, and to Diane Kucharczyk for her lab experience. I would also like to thank all those who volunteered to be subjects for the experiments, and those who tolerated the sometimes odd intrusions into the hallways and bathrooms. Thanks to Jeff Pelz, Eriko Miyahara, and Mary Hayhoe for their thoughtful insights and recommendations. Finally, thanks to my family for understanding what the reason was for those many late nights and missed events.
Table of Contents

List of Figures ix

List of Tables xviii

1. Introduction 1
   1.1 Overview ................................................................. 1
   1.2 Objectives (Statement of Work) ..................................... 5

2. Background 6
   2.1 Historical Perspective ................................................. 6
   2.2 Eye Movements ......................................................... 12
   2.3 Visual Attention and Selection ...................................... 16
       2.3.1 Saliency Maps ..................................................... 16
       2.3.2 Feature Integration Theory of Attention ..................... 17
   2.4 The World as Anchor .................................................. 18
       2.4.1 Semantic Consistency ............................................. 18
       2.4.2 Change Blindness ................................................. 18
       2.4.3 Exocentric Reference Frames .................................. 19
       2.4.4 Position Constancy During Passive Movement ............. 19
   2.5 Perceiving the Direction of Heading During Motion .......... 20
       2.5.1 Retinal vs. Extra-retinal Information ....................... 20
       2.5.2 Differential Motion Parallax .................................. 23
   2.6 The Effects of Freeing the Head ..................................... 23
   2.7 Natural Tasks .......................................................... 26
       2.7.1 Memory Representation of a Simple Natural Task - Blocks Copying .... 27
       2.7.2 Sequential Looking Task - Tapping vs. Looking Only ........ 29
       2.7.3 Visual Memory in Problem Solving - Geometry ................ 31
       2.7.4 Eye Movements and Work Load During Driving .............. 32
       2.7.5 The Direction of Gaze During Driving ....................... 34
       2.7.6 Eye Movements While Making Tea ................................ 35
       2.7.7 Portable Eye-Tracking Natural Tasks ......................... 38
3. Approach 41
  3.1 History of Eye-tracking Methods 42
  3.1.1 Electrical Methods 42
  3.1.2 Optical Methods 43
  3.2 The VPL Portable, Wearable Eye-Tracker 45
  3.2.1 The Custom Goggles Headgear 46
  3.2.2 Other System Components 50
  3.2.3 Theory of Operation 52
  3.2.4 Eye-Tracker Set-Up and Calibration 54
  3.2.5 Eye Movement Monitoring 56
  3.3 Example of Real-Time Data Capture 57
  3.4 Data Analysis 59
    3.4.1 Coding the Data 60
    3.4.2 Fixation Durations and Saccade Size 61
    3.4.3 Statistical Analysis 63
  4. Experiment 1 - Moving Through a Hallway 66
    4.1 Objective 67
    4.2 Experimental Design and Conditions 68
    4.3 Data Analysis and Results 71
    4.4 Conclusion 88
  5. Experiment 2 - Fixation Stability 89
    5.1 Objective 89
    5.2 Experimental Design and Conditions 90
    5.3 Data Analysis and Results 93
    5.4 Conclusion 99
  6. Experiment 3 - Handwashing 100
    6.1 Objective 100
    6.2 Experimental Design and Conditions 100
    6.3 Data Analysis and Results 101
      6.3.1 Fixation Durations 101
      6.3.2 Saccade Size 104
      6.3.3 Major Sub-tasks 107
      6.3.4 "Look-aheads" 110
    6.4 Conclusion 115
  7. Experiment 4 - Making a Selection From a Vending Machine 116
    7.1 Objective 116
    7.2 Experimental Design and Conditions 117
    7.3 Data Analysis and Results 119
      7.3.1 Fixation Durations 119
      7.3.2 Saccade Size 121
      7.3.3 Comparison of Handwashing and Vending Machine Experiments 123
      7.3.4 Major Sub-tasks 124
List of Figures

Figure 2-1  Humans have a perceptual bias toward seeing the triangle as a whole ..........7
Figure 2-2  Scan paths are tasks dependent ..........................................................10
Figure 2-3  The distribution of rods and cones is unevenly distributed across the retina...13
Figure 2-4  Vergence eye movement ......................................................................15
Figure 2-5  Experimental conditions for sufficiency of retinal information ..............21
Figure 2-6  The effect of freeing the head on fixation stability .................................25
Figure 2-7  The effects of freeing the head on saccades .........................................26
Figure 2-8  Blocks copying task .............................................................................27
Figure 2-9  Eye movement strategies used for blocks copying task........................28
Figure 2-10 Tapping vs. Looking only .................................................................30
Figure 2-11 Fixation durations for expert and novice drivers, for several conditions ......33
Figure 2-12 Fixation duration as a function of task difficulty for a driving task ........34
Figure 3-1  Portable, wearable eye-tracking headgear ...........................................46
Figure 3-2  Optics module (HMO) of headgear .....................................................47
Figure 3-3  Top view of headgear ...........................................................................47
Figure 3-4  Eye-tracking hardware, not including the goggles ...............................50
Figure 3-5  Camcorder inside backpack ..........................................................50
Figure 3-6  Close up front panel of control unit ...........................................50
Figure 3-7  Subject wearing eye-tracking gear, ready to perform an experiment ....51
Figure 3-8  Image of pupil (white outline) and corneal reflection (black outline) ........52
Figure 3-9  Calculation of the line of gaze .......................................................53
Figure 3-10 Diffraction pattern used for calibration ........................................55
Figure 3-11 Real Data Capture .......................................................................57
Figure 3-12 Trace of vertical eye position .......................................................58
Figure 3-13 Trace of horizontal eye position ....................................................58
Figure 3-14 Expanded view of vertical eye position .........................................58
Figure 3-15 Expanded view of horizontal eye position .....................................58
Figure 3-16 Eye-tracker noise, no averaging ....................................................59
Figure 3-17 Eye-tracker noise, 2 field averaging .............................................59
Figure 3-18 Eye-tracker noise, 4 field averaging .............................................59
Figure 3-19 Eye-tracker noise, 8 field averaging .............................................59
Figure 3-20 Calculation of visual angle from field of view ...............................62
Figure 3-21 The gamma density function with B = 1 and A = 2 .......................65
Figure 4-1  Simulated condition set-up for Exp. 1 ..........................................68
Figure 4-2  First hallway for Exp. 1 .................................................................69
Figure 4-3  Second hallway for Exp. 1 ............................................................69
Figure 4-4  Third hallway for Exp. 1 ...............................................................69
Figure 4-5  Fourth hallway for Exp. 1 ..............................................................69
Figure 4-6  Simulated/Active ...........................................................................70
Figure 4-7  Simulated/Passive .........................................................................70
Figure 4-8  Real/Active ..........................................................70
Figure 4-9  Real/Passive ..........................................................70
Figure 4-10 Fixation Durations - Simulated/Active - JB .........................71
Figure 4-11 Fixation Durations - Simulated/Active - JP ..........................71
Figure 4-12 Fixation Durations - Simulated/Active - MA ..........................71
Figure 4-13 Fixation Durations - Simulated/Active - RC ..........................71
Figure 4-14 Fixation Durations - Simulated/Passive - JB .........................72
Figure 4-15 Fixation Durations - Simulated/Passive - JP ..........................72
Figure 4-16 Fixation Durations - Simulated/Passive - MA ..........................72
Figure 4-17 Fixation Durations - Simulated/Passive - RC ..........................72
Figure 4-18 Fixation Durations - Real/Active - JB ................................73
Figure 4-19 Fixation Durations - Real/Active - JP ................................73
Figure 4-20 Fixation Durations - Real/Active - MA ................................73
Figure 4-21 Fixation Durations - Real/Active - RC ................................73
Figure 4-22 Fixation Durations - Real/Passive - JB ................................74
Figure 4-23 Fixation Durations - Real/Passive - JP ................................74
Figure 4-24 Fixation Durations - Real/Passive - MA ................................74
Figure 4-25 Fixation Durations - Real/Passive - RC ................................74
Figure 4-26 Fixation Durations - Strings of Length x - Simulated/Active - JB .........77
Figure 4-27 Fixation Durations - Strings of Length x - Simulated/Passive - JB .........77
Figure 4-28 Fixation Durations - Strings of Length x - Simulated/Active - JP .........77
Figure 4-29 Fixation Durations - Strings of Length x - Simulated/Passive - JP .........77
Figure 4-30 Fixation Durations - Strings of Length x - Simulated/Active - MA .........77
Figure 4-31 Fixation Durations - Strings of Length x - Simulated/Passive - MA .........77
Figure 4-32  Fixation Durations - Strings of Length x - Simulated/Active - RC ..........78
Figure 4-33  Fixation Durations - Strings of Length x - Simulated/Passive - RC ..........78
Figure 4-34  Fixation Durations - Strings of Length x - Real/Active - JB ...................78
Figure 4-35  Fixation Durations - Strings of Length x - Real/Passive - JB ...................78
Figure 4-36  Fixation Durations - Strings of Length x - Real/Active - JP ...................78
Figure 4-37  Fixation Durations - Strings of Length x - Real/Passive - JP ...................78
Figure 4-38  Fixation Durations - Strings of Length x - Real/Active - MA ...................78
Figure 4-39  Fixation Durations - Strings of Length x - Real/Passive - MA ...................78
Figure 4-40  Fixation Durations - Strings of Length x - Real/Active - RC ...................78
Figure 4-41  Fixation Durations - Strings of Length x - Real/Passive - RC ...................78
Figure 4-42  Saccade Size - Simulated/Active - JB ..............................................80
Figure 4-43  Saccade Size - Simulated/Active - JP ..............................................80
Figure 4-44  Saccade Size - Simulated/Active - MA ..............................................80
Figure 4-45  Saccade Size - Simulated/Active - RC ..............................................80
Figure 4-46  Saccade Size - Simulated/Passive - JB ..............................................81
Figure 4-47  Saccade Size - Simulated/Passive - JP ..............................................81
Figure 4-48  Saccade Size - Simulated/Passive - MA ..............................................81
Figure 4-49  Saccade Size - Simulated/Passive - RC ..............................................81
Figure 4-50  Saccade Size - Real/Active - JB ..............................................82
Figure 4-51  Saccade Size - Real/Active - JP ..............................................82
Figure 4-52  Saccade Size - Real/Active - MA ..............................................82
Figure 4-53  Saccade Size - Real/Active - RC ..............................................82
Figure 4-54  Saccade Size - Real/Passive - JB ..............................................83
Figure 4-55  Saccade Size - Real/Passive - JP ..............................................83
Figure 4-56 Saccade Size - Real/Passive - MA ........................................83
Figure 4-57 Saccade Size - Real/Passive - RC ........................................83
Figure 4-58 Fixation Duration vs. Saccade Size - Simulated/Active - JB ..............86
Figure 4-59 Fixation Duration vs. Saccade Size - Simulated/Passive - JB ..............86
Figure 4-60 Fixation Duration vs. Saccade Size - Simulated/Active - JP ..............86
Figure 4-61 Fixation Duration vs. Saccade Size - Simulated/Passive - JP ..............86
Figure 4-62 Fixation Duration vs. Saccade Size - Simulated/Active - MA ..............86
Figure 4-63 Fixation Duration vs. Saccade Size - Simulated/Passive - MA ..............86
Figure 4-64 Fixation Duration vs. Saccade Size - Simulated/Active - RC ..............86
Figure 4-65 Fixation Duration vs. Saccade Size - Simulated/Passive - RC ..............86
Figure 4-66 Fixation Duration vs. Saccade Size - Real/Active - JB .....................87
Figure 4-67 Fixation Duration vs. Saccade Size - Real/Passive - JB .....................87
Figure 4-68 Fixation Duration vs. Saccade Size - Real/Active - JP .....................87
Figure 4-69 Fixation Duration vs. Saccade Size - Real/Passive - JP .....................87
Figure 4-70 Fixation Duration vs. Saccade Size - Real/Active - MA .....................87
Figure 4-71 Fixation Duration vs. Saccade Size - Real/Passive - MA .....................87
Figure 4-72 Fixation Duration vs. Saccade Size - Real/Active - RC .....................87
Figure 4-73 Fixation Duration vs. Saccade Size - Real/Passive - RC .....................87
Figure 5-1 Set-up for Condition 1: Small Screen ..........................................91
Figure 5-2 Set-up for Condition 2: Large Screen ..........................................92
Figure 5-3 Set-up for Condition 3: Real Walking ..........................................92
Figure 5-4 Fixation stability for target on right for subject JB .........................94
Figure 5-5 Fixation stability for target straight ahead for subject JB ...................94
Figure 5-6 Fixation stability for target on right for subject JK .........................96
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-7</td>
<td>Fixation stability for target straight ahead for subject JK</td>
<td>96</td>
</tr>
<tr>
<td>5-8</td>
<td>Fixation stability for small screen condition</td>
<td>97</td>
</tr>
<tr>
<td>5-9</td>
<td>Fixation stability for large screen condition</td>
<td>97</td>
</tr>
<tr>
<td>5-10</td>
<td>Fixation stability for real walking condition</td>
<td>97</td>
</tr>
<tr>
<td>6-1</td>
<td>Men's washroom used for Exp. 3</td>
<td>101</td>
</tr>
<tr>
<td>6-2</td>
<td>Women's washroom used for Exp. 3</td>
<td>101</td>
</tr>
<tr>
<td>6-3</td>
<td>Fixation Durations - Handwashing - Subject AS</td>
<td>102</td>
</tr>
<tr>
<td>6-4</td>
<td>Fixation Durations - Handwashing - Subject JB</td>
<td>102</td>
</tr>
<tr>
<td>6-5</td>
<td>Fixation Durations - Handwashing - Subject JP</td>
<td>102</td>
</tr>
<tr>
<td>6-6</td>
<td>Fixation Durations - Handwashing - Subject MH</td>
<td>102</td>
</tr>
<tr>
<td>6-7</td>
<td>Fixation Durations - Not Including Wash Hands - Subject AS</td>
<td>103</td>
</tr>
<tr>
<td>6-8</td>
<td>Fixation Durations - Not Including Wash Hands - Subject JB</td>
<td>103</td>
</tr>
<tr>
<td>6-9</td>
<td>Fixation Durations - Not Including Wash Hands - Subject JP</td>
<td>103</td>
</tr>
<tr>
<td>6-10</td>
<td>Fixation Durations - Not Including Wash Hands - Subject MH</td>
<td>103</td>
</tr>
<tr>
<td>6-11</td>
<td>Saccade Size - Handwashing - Subject AS</td>
<td>105</td>
</tr>
<tr>
<td>6-12</td>
<td>Saccade Size - Handwashing - Subject JB</td>
<td>105</td>
</tr>
<tr>
<td>6-13</td>
<td>Saccade Size - Handwashing - Subject JP</td>
<td>105</td>
</tr>
<tr>
<td>6-14</td>
<td>Saccade Size - Handwashing - Subject MH</td>
<td>105</td>
</tr>
<tr>
<td>6-15</td>
<td>Saccade Size - Not Including Wash Hands - Subject AS</td>
<td>106</td>
</tr>
<tr>
<td>6-16</td>
<td>Saccade Size - Not Including Wash Hands - Subject JB</td>
<td>106</td>
</tr>
<tr>
<td>6-17</td>
<td>Saccade Size - Not Including Wash Hands - Subject JP</td>
<td>106</td>
</tr>
<tr>
<td>6-18</td>
<td>Saccade Size - Not Including Wash Hands - Subject MH</td>
<td>106</td>
</tr>
<tr>
<td>6-19</td>
<td>Handwashing - Subject AS</td>
<td>108</td>
</tr>
<tr>
<td>6-20</td>
<td>Handwashing - Subject JB</td>
<td>108</td>
</tr>
</tbody>
</table>
Figure 6-21  Handwashing - Subject JP  ......................................................... 108
Figure 6-22  Handwashing - Subject MH  ......................................................... 108
Figure 6-23  Handwashing - Subject AS  ......................................................... 109
Figure 6-24  Handwashing - Subject JB  ......................................................... 109
Figure 6-25  Handwashing - Subject JP  ......................................................... 109
Figure 6-26  Handwashing - Subject MH  ......................................................... 109
Figure 6-27  Elapsed time between object fixation and object manipulation for
             Handwashing - Subject AS  ......................................................... 111
Figure 6-28  Elapsed time between object fixation and object manipulation for
             Handwashing - Subject JB  ......................................................... 111
Figure 6-29  Elapsed time between object fixation and object manipulation for
             Handwashing - Subject JP  ......................................................... 111
Figure 6-30  Elapsed time between object fixation and object manipulation for
             Handwashing - Subject MH  ......................................................... 111
Figure 6-31  JB’s first fixation on garbage can  ............................................. 112
Figure 6-32  JB’s first manipulation of garbage can, 7.1 seconds later  .......... 112
Figure 6-33  Handwashing - Subject AS  ......................................................... 114
Figure 6-34  Handwashing - Subject JB  ......................................................... 114
Figure 6-35  Handwashing - Subject JP  ......................................................... 114
Figure 6-36  Handwashing - Subject MH  ......................................................... 114
Figure 7-1  The alcove with the four vending machines used in Experiment 4 .......... 117
Figure 7-2  Coffee machine ................................................................. 118
Figure 7-3  Soda machine ................................................................. 118
Figure 7-4  Candy/chip machine ............................................................ 118
Figure 7-5  Sandwich machine ............................................................. 118
Figure 7-6  Fixation Durations - Vending Machine - Subject AS  .................... 120
Figure 7-7  Fixation Durations - Vending Machine - Subject JB ..........................120
Figure 7-8  Fixation Durations  Vending Machine - Subject JT ..........................120
Figure 7-9  Fixation Durations  Vending Machine - Subject MH ..........................120
Figure 7-10 Saccade Size - Vending Machine - Subject AS ..............................122
Figure 7-11 Saccade Size - Vending Machine - Subject JB ..............................122
Figure 7-12 Saccade Size - Vending Machine - Subject JT ..............................122
Figure 7-13 Saccade Size - Vending Machine - Subject MH ..............................122
Figure 7-14 Vending Machine - Soda Machine - Subject AS .............................124
Figure 7-15 Vending Machine - Sandwich Machine - Subject JB ........................124
Figure 7-16 Vending Machine - Candy/Chip Machine - Subject JT ........................124
Figure 7-17 Vending Machine - Coffee Machine - Subject MH ...........................124
Figure 7-18 Vending Machine - Subject AS .....................................................126
Figure 7-19 Vending Machine - Subject JB .....................................................126
Figure 7-20 Vending Machine - Subject JT .....................................................126
Figure 7-21 Vending Machine - Subject MH .....................................................126
Figure 7-22 Elapsed time between object fixation and object manipulation for Vending Machine - Subject AS ..........................128
Figure 7-23 Elapsed time between object fixation and object manipulation for Vending Machine - Subject JB ..........................128
Figure 7-24 Elapsed time between object fixation and object manipulation for Vending Machine - Subject JT ..........................128
Figure 7-25 Elapsed time between object fixation and object manipulation for Vending Machine - Subject MH ..........................128
Figure 7-26 Vending Machine - Subject AS .....................................................130
Figure 7-27 Vending Machine - Subject JB .....................................................130
Figure 7-28 Vending Machine - Subject JT .....................................................130
Figure 7-29  Vending Machine - Subject MH ..................................................130
Figure 7-30  JB looks at exit of candy/chip machine at timecode of 00:08:05:02 ........132
Figure 7-31  JB retrieves purchase from machine at timecode of 00:08:11:26, 6.8 seconds later .................................................................132
Figure 7-32  MH looks at coffee cup lids at a timecode of 00:10:03:27 .................132
Figure 7-33  MH looks at coffee cup lids at a timecode of 00:10:40:10 ..................132
Figure 7-34  MH retrieves coffee cup lid at a timecode of 00:10:56:27, 53 seconds after first locating the lids, and 16.5 seconds after a second look to the lids ......132
Figure 7-35  Fixation Durations and Subtasks - Vending Machine - Subject AS ........134
Figure 7-36  Fixation Durations and Subtasks - Vending Machine - Subject JB ..........134
Figure 7-37  Fixation Durations and Subtasks - Vending Machine - Subject JT ........134
Figure 7-38  Fixation Durations and Subtasks - Vending Machine - Subject MH ..........134
Figure 7-39  Fixation Durations and Subtasks - Handwashing - Subject AS .............135
Figure 7-40  Fixation Durations and Subtasks - Handwashing - Subject JB ..............135
Figure 7-41  Fixation Durations and Subtasks - Handwashing - Subject JP ..............135
Figure 7-42  Fixation Durations and Subtasks - Handwashing - Subject MH .............135
Figure 7-43  Saccade Size and Subtasks - Vending Machine - Subject AS .................136
Figure 7-44  Saccade Size and Subtasks - Vending Machine - Subject JB .................136
Figure 7-45  Saccade Size and Subtasks - Vending Machine - Subject JT ..................136
Figure 7-46  Saccade Size and Subtasks - Vending Machine - Subject MH .................136
Figure 7-47  Saccade Size and Subtasks - Handwashing - Subject AS .....................137
Figure 7-48  Saccade Size and Subtasks - Handwashing - Subject JB .....................137
Figure 7-49  Saccade Size and Subtasks - Handwashing - Subject JP .....................137
Figure 7-50  Saccade Size and Subtasks - Handwashing - Subject MH .....................137
List of Tables

Table 2-1  Some computations can be simplified by making assumptions about behavior .................................................................9
Table 3-1  Example of code from eye-tracking experiment .........................................................61
Table 7-1  Comparison of fixation durations for handwashing and vending machine ....123
Table 7-2  Comparison of saccade size for handwashing and vending machine ........123
Chapter 1

1. Introduction

1.1 Overview

The purpose of vision is to serve the needs of the individual. As an individual goes about performing day-to-day activities, the visual system is continually monitoring the environment to provide information about each interaction; information that enables meaningful interactions with that environment for the fulfillment of a plan of action.

In this sense, vision is not a passive process whereby information is merely collected, processed, and stored for later retrieval, but rather an active process that integrates goal-oriented behavior with proprioceptive signals from the individual's physical state, and exteroceptive information about the layout of the environment.

Visual perception is essentially a selective process. The particular sequence of selections is largely dependent upon the task to be performed, and as such is driven by the goals of the individual, but each discrete selection occurs mostly at a subconscious level. Eye movements are one of the mechanisms by which the selection process proceeds. The human eye can be thought of as a space-variant sensor that exhibits non-uniform sampling
across the retina. A central, high resolution fovea encompassing approximately 1° of visual angle is surrounded by a low resolution periphery. An eye movement is required to bring an object of interest to the fovea, and is the means for sustaining overt attention on the object. The apparent purpose of eye movements thus appears to be that of allowing for the impression of a broad, high resolution visual field from multiple sequential fixations. This observation, based on the verifiable physiology of the human eye, does not adequately offer an answer to a central question regarding the role of eye movements in visual perception: where is attention to be focused next?

This research effort is largely concerned with providing a framework within which that question may be approached. There is obviously not a single answer that will apply to every situation requiring focused visual attention, but it is possible to extract a certain amount of commonality in everyday tasks that gives rise to particular patterns of selection.

A primary objective of this research is to study eye movements of subjects while they perform everyday tasks in a natural environment. Much of what is currently understood about human eye movements, and also about visual perception in general, is based on psychophysical studies conducted in the confines of a laboratory setting. Since humans did not evolve their sensory-perceptual abilities in such a restricted environment, it is valid to question whether or not the results obtained from such studies apply in a practical sense. It is also possible that subjects may exhibit an unconscious bias while performing in a laboratory, providing results that are valid in the laboratory, but not necessarily so in the world outside of it.

The psychophysical "black-box" approach to studying eye movements in isolation, and not in the context of a rich, interactive environment suffers from other methodological flaws. One is the linear-systems approach, which assumes that an outcome of an event can
be approximated by the linear sum of each input acting separately. In this view, if a single input can be isolated from all other possible inputs, its effect on the outcome can be precisely measured and quantified. All such inputs, when taken together, describe the system response. In this context, linearity refers to the idea that the whole is equal to the sum of its parts. The assumption of linearity as applied to human eye movements has not been adequately shown to be valid. A secondary goal of this research project is to provide grounds either for or against that assumption. The means for doing this is provided by the RIT portable, wearable eye-tracker, which was developed for the purpose of studying subjects' eye movements while they are performing common, everyday tasks in a natural, unrestricted environment.

A portable, head mounted eye-tracker was used for this research, as well as hardware and software that enabled a computation of the line-of-gaze for a subject who is wearing the eye-tracker and performing tasks in a natural environment. The line-of-gaze is displayed as a cursor superimposed on a video scene of the environment as seen by the subject. Data analysis of the cursor position as a function of time correlates to eye movements and affords an indirect method of determining the cognitive processes underlying visual perception.

A final objective of this research was to consider the implications of a sequential fixation strategy and of non-uniform sampling, or foveation, for an artificial vision system. Researchers in robotics and computer vision often consider the human visual system as a model for artificial vision systems. It would be beneficial to be able to describe the high-level cognitive processes underlying visual perception in a way that would be amenable to a computer program. Active computer vision is an area of current research, and much work needs to be done to be able to develop and implement efficient algorithms. Eye-tracking
human subjects while they are engaged in natural tasks may offer insights into how to design those algorithms.

In summary, the objectives of this research project are three-fold: to consider the effects of freeing the subject from the restraints of the laboratory setting during eye-tracking experiments, to develop a framework for describing the temporal sequencing of fixations across tasks as well as within tasks, and to evaluate the appropriateness of such a framework for serving as a model for an artificial vision system. Since the purpose of vision is to serve the needs of the individual, it seems reasonable to conclude that a hypothesis about where to look next can best be formulated by considering the cooperative relationship between vision and action.
1.2 Objectives (Statement of Work)

Following are the objectives mandated for this research:

a) Conduct a literature review of the subject. The topics related to the topic are: eye movements, visuo-motor coordination, selective attention, plan schemata, active vision, and animate vision.

b) Design a series of experiments to monitor subjects' eye movements as they perform a range of common, everyday tasks selected to gain an understanding of the interaction between vision and action. Such tasks include:
   i) Walking along a corridor, being pushed in a wheelchair, and watching a videotape of someone walking along a corridor or being pushed in a wheelchair
   ii) Maintaining fixation on an object while walking along a corridor
   iii) Washing one's hands in a lavatory
   iv) Making a selection from a vending machine

c) Recruit subjects and carry out the experimentation.

d) Analyze the data collected in terms of eye movement metrics. Examples of such metrics are: fixation duration, number of fixations, saccade length, saccades per second, etc.

e) Study the results to determine the pre-conscious strategies used by individuals as they performed the tasks.

f) Modify and repeat the experimentation and data analysis as necessary to investigate any interesting or emergent pattern of oculomotor behavior.

g) Formulate conclusions based on results. Demonstrate the usefulness of results, and show how they can be applied to artificial active vision systems.
Chapter 2

2. Background

2.1 Historical Perspective

In 1867 Herman von Helmholtz published his thoughts on the nature of visual perception in a book entitled *Treatise on Physiological Optics* (Helmholtz, 1867/1925). This work laid the foundation for the classical approach to the philosophical treatment of vision known as constructivism. The goal of constructivism was to explain visual perception as arising from the confluence of many local information processing units, which when combined together, construct a global percept of the world. A central tenet of modern constructivism is the belief that perception relies upon a process of unconscious inference. In other words, in order for local information to be bound with other local information in a meaningful way, an inference must be made about the most likely interpretation.

An example of how unconscious inference could be used to explain perception is shown in Figure 2-1. Two possible interpretations of image A are shown as image B and image C. Since there is insufficient information available in image A to decide upon the
correct interpretation, a constructivist would invoke the principle of unconscious inference to explain the human perceptual bias of choosing image B as the correct interpretation. Image B is chosen because it is the most likely possibility.

![Image 2-1](image.png)

**Figure 2-1. Humans have a perceptual bias toward seeing the triangle as whole.**

The inference is largely unconscious in that the observer is generally not aware that probabilities are being compared, and that logical inferences are being made.

A constructivist approach to the inverse problem – that is, the problem of how 2-D retinal images are transformed into a perception of the 3-D environment – would be to consider the 2-D retinal image as belonging to the most likely state of affairs in the environment that would give rise to such an image.

In contrast to the constructivist theory of unconscious inference, an ecological perspective was espoused by James Gibson (Gibson, 1966), who argued that direct perception of the environment is sufficient for solving the inverse problem. He believed that all visual perception is the result of the interaction between the observer and surfaces, or more specifically the light reflected off surfaces, in the environment. Surfaces are composed of texture elements, and it is the structure that exists in the surfaces that in turn structures the light that reaches the eye of the observer. When the observer moves around the surfaces, the changing ambient optic array of light reaching the retina results in an optic flow field that is unique for every point in the environment.
Thus, the inverse problem is solved by considering the movement of the observer as integral to the reconstruction. Change in structure over time supplies the missing dimension.

In the late 1970's David Marr (Marr, 1982) combined the theoretical constructs from both constructivism and ecological perception to create the first computational approach for describing visual processes. He used mathematical techniques to develop computer programs that simulated biological vision, and led the early efforts of computational and computer scientists who designed the first machine vision systems.

Marr disagreed with Gibson, however, on the issue of representation. Gibson held that the environment is the repository for all of the information that is necessary for visual interaction, whereas Marr believed that the external world is represented internally, in all of its detail. An example of the internal representation is what Marr calls the "2½ dimension" sketch, an internal retinotopic image with the potential for a 3-D representation.

Marr's work has had a strong influence on the current understanding of early vision, and this understanding has led to a number of computational approaches based on early, or low-level biological vision. It is assumed that in order to simulate a process as complex as high-level visual perception, one must begin with, and correctly implement, the lower level processes. Only then will the "correct" way to implement the higher-level cognitive processes become apparent.

Ballard and Brown pointed out several weaknesses to this approach (Ballard & Brown, 1992). First, early visual processes do not take into account the motivation of the observer. Marr's treatment of the visual process as purely passive precludes a potentially rich source of information that may help to constrain the inverse problem,
and may make high-level computational vision more feasible. It may be helpful to think of cognition as the driving force behind the collection of low-level information, instead of thinking of it as merely the result of a collection of responses.

Second, the early vision approach does not take into account sequentialization and gaze control that humans use to make efficient use of the multi-resolution capabilities of the human eye. Finally, Marr's model does not make use of learning strategies or adaptational responses to the environment. His model of perception is essentially a rich, highly detailed, task independent description of the world, which is continually being called upon by cognition for performing specific tasks. Ballard and Brown (1992) describe an alternate way of approaching the complexities imposed by vision, and suggest numerous simplifications that would result from taking behavioral assumptions into account. Their findings, which are exemplified by a construct called animate vision, are summarized in Table 2-1 below.

### Computations Simplified by Behavioral Assumptions

<table>
<thead>
<tr>
<th>Agent's Behavior</th>
<th>Behavioral Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape from shading</td>
<td>Light source not directly behind viewer</td>
</tr>
<tr>
<td>Time to adjacency</td>
<td>Rectilinear motion; gaze in the direction of motion</td>
</tr>
<tr>
<td>Kinetic depth</td>
<td>Lateral head motion while fixating a point in a stationary world</td>
</tr>
<tr>
<td>Color homing</td>
<td>Target object is distinguished by its color spectrum</td>
</tr>
<tr>
<td>Optic flow</td>
<td>Texture-rich environment</td>
</tr>
<tr>
<td>Stereo depth</td>
<td>System can fixate environmental points</td>
</tr>
<tr>
<td>Edge homing</td>
<td>Target position can be described by approximate directions from texture in its surround</td>
</tr>
<tr>
<td>Object tracking</td>
<td>Vergence can be used to improve tracking performance</td>
</tr>
</tbody>
</table>

Table 2-1. Some computations can be simplified by making assumptions about behavior. From Ballard & Brown, 1992.
Another objection to the early-vision approach toward computational vision is suggested by the work conducted by Yarbus in the 1960's. Yarbus showed how high-level cognitive events are reflected in the patterns of eye-movement traces (Yarbus, 1967). He found that different patterns of eye-movement traces, or scan-paths, could be elicited from subjects when they performed context-sensitive tasks. For example, when subjects were shown a painting depicting a scene of several people greeting an unexpected visitor, a specific question posed to the subjects elicited a specific "signature" pattern of eye movements. Different questions elicited different "signature" patterns. Figure 2-2 below shows the painting and typical scan paths for a subject formulating an answer to the various questions.

```
Figure 2-2. Scan paths are task dependent. From Palmer, 1999 and Yarbus, 1967.
```

The observation that oculomotor behavior is largely task dependent leads one to consider how other aspects of visual perception may be dependent upon the cognitive goals
of the observer. David Lee has suggested that information processing by humans should be considered in the context of a unified perceptuo-motor system, which is itself a part of the organism-environment system (Lee, 1978, 1980). His ideas pertaining to the functions of vision are an extension of the ecological perceptual model set forth by Gibson a decade earlier. In his view, the human visual system must be studied not only in an environmental context, but also in the context of the individual’s sensory-motor system. Vision is functionally inseparable from the motor system. Information becomes available to the individual via three separate sources: exteroceptive, proprioceptive, and exproprioceptive. The exteroceptive source delivers information about the layout and affordances of the environment. The proprioceptive source delivers information about the position, orientation, and movement of the body or parts of the body. The exproprioceptive source delivers information about the union of the exteroceptive and proprioceptive sources, information about the movement of the body relative to the environment. The exproprioceptive information represents the interaction between the individual and space over time.

Taken together, the three sources provide the means for a cooperative relationship to exist between vision and action. Goal-oriented behavior, planning, and decision-making all play a significant part in the visual perception experienced by the individual.

To summarize the history of formal thinking about the nature of human visual perception, the constructivist and computational early-vision approaches taken by Helmholtz and Marr emphasize the autonomy of the individual and unconscious mechanisms to guide the visual perceptual process. This is the foundation for much of the current linear-systems methods for teasing apart the factors that influence perception. Gibson, Lee, and Ballard & Brown, on the other hand think of visual perception in terms of
the interaction between the individual and the environment according to goals, actions, motivation and behavior. For them, trying to understand how vision works by studying subjects' responses to artificial stimuli in a laboratory setting is like trying to understand how fish swim by putting them in a sandbox. From this point of view then, the factors that have had a major evolutionary influence on vision and that have largely shaped human visual perception are precisely those factors that are missing from the laboratory setting.

2.2 Eye Movements

The binocular visual field subtends an area approximately 130° vertically and 180° horizontally. Most of that area contains low-resolution peripheral information. In order to obtain detailed, high-resolution information from different areas in the environment, the eyes must move. The purpose of an eye movement is to bring the most visually relevant part of a scene onto the area of the retina with the highest visual acuity, and to keep it there during focused attention. This area is called the fovea and subtends approximately one degree of visual angle, covering an area of the visual field approximately equal to the size of a thumbnail extended at an arm’s length. Attention can then be re-deployed to another area in the visual field to initiate the next eye movement.

The photoreceptors of the human eye consist of both rods and cones, the cones being the photoreceptors responsible for color perception and visual acuity. As shown in Figure 2-3, the population of cones is highest in the fovea and falls off rapidly toward the periphery. There is a 1:1 or greater correspondence between photoreceptors and ganglion cells in the fovea, however this ratio increases continuously along the periphery. This fact, along with the higher concentration of cones in the fovea, accounts for the higher visual acuity there. Figure 2-3 is an illustration of the distribution of the rods and cones across the retina.
Figure 2-3. The distribution of rods and cones is unevenly distributed across the retina. The fovea contains the highest concentration of cones, for high visual acuity in that region. From Palmer, 1998.

Traditionally, eye movements have been classified into six categories:

1. Miniature eye movements - These are the only type of eye movements that do not have a selective function. They include tremors in the extraocular muscles that control rotation of the eyes in their spherical socket, drift of the foveated image, and microsaccades to bring the drifted image back to the fovea. The result is constant motion of the optical image on the retina.

2. Saccades - These are high velocity, ballistic eye movements that have the function of bringing images of objects of interest to the fovea. It is generally believed that once a saccadic eye movement has begun, it cannot be altered. A typical saccade takes approximately 150 - 200 msec to plan and execute; planning takes about 150 msec on average, and the duration of the eye movement is approximately 20 msec plus 2 msec per degree of visual angle (Carpenter, 1988). Saccades can reach velocities up to 600° per second, and individuals typically make 3 or 4 saccades per second, depending on the task being performed.
Studies on eye movements during reading have shown that saccades during reading are typically seven letters long, which is a saccade length of between 1° and 2° for reading standard size text at a distance of 40 cm (O'Reagan, 1990). It has also been found that there is a wide distribution of within-word target landing for reading text. In other words, there is no precise position within the word that the eye targets the saccade to land on, anywhere within the word is sufficient for comprehension (Morgan, et. al., 1990). Fixations are defined as the time between successive saccades; a typical fixation duration for reading is between 200 and 300 msec.

3. Smooth pursuit - These eye movements track the position of a moving object, with the purpose of keeping the image in the foveal region. Ideally, the image remains stationary on the retina. After an initial saccade to track the moving object, the eye movement is smooth and continuous, as opposed to the abruptness of saccades. Constant correction of image position on the fovea is maintained by means of a feedback signal from the brain that senses the position of the object as it moves. Thus smooth pursuit cannot usually be maintained in the absence of a moving target. The maximum velocity is approximately 100° per second; target velocities higher than that cause retinal slippage and disable the tracking mechanism. During pursuit, the image of the pursued object is clear, with all other untracked objects smeared due to their relative motion on the retina.

4. Vergence - When an observer fixates an object, the eyes converge toward one another, with the degree of convergence depending upon the distance between the observer and the object. Vergence eye movements are disconjugate in the sense that the eyes rotate opposite to one another. For a conjugate movement such as pursuit, the eyes rotate in the same direction. If an object is moving both in depth and in direction, a disconjugate and a conjugate component of the eye movement is necessary to track it.
5. Vestibular - When the head rotates, the vestibular ocular reflex (VOR) allows us to fixate an object in the environment without visual feedback. The information necessary to control eye movements when the head moves originates in the vestibular system of the inner ear, which senses the orientation of the head. Vestibular eye movements are faster than pursuit movements, however, high velocity head movements such as those encountered while running or walking fast cannot be fully compensated for by a vestibular eye movement (Palmer, 1999). When this happens, objects in the environment that require high visual acuity for perception (such as lettering on signs) will appear blurred.

6. Optokinetic - An optokinetic eye movement is a response to the rapid translation of the entire visual field, or a large part of it. For example, if an observer is looking through a window at a train passing by, fixating and tracking a spot on the train will cause the observer to exhibit the optokinetic response. It is characterized by a slow, tracking phase in which the image is stabilized on the retina, followed by a rapid, saccade-like snap of the eyes in the direction opposite to the image motion. This is known as optokinetic nystagmus (OKN) and causes the image to blur.
Recent studies have suggested that there are actually only two categories of eye movements: saccadic and smooth (Steinman, Kowler, and Collewijn, 1990). The claim is that the classification into six categories is artificial, a result of the early laboratory methods that studied simple tasks in a constrained and sparse visual environment. The experimental results of such early studies reflected the low-level and involuntary aspects of oculomotor control, and were simply responses to sensory cues that did not reflect the cognitive processes that people typically employ while engaged in natural tasks such as expectation, motivation, and learning.

2.3 Visual Attention and Selection

The mechanics of oculomotor behavior do not explain how the selection process is controlled. Questions such as "what is the region of interest?", and "where should the next fixation be?" can best be answered within a framework that considers the purpose of focused attention.

2.3.1 Saliency Maps

The notion of a saliency map was proposed to define the relationship between the components of a scene according to their relative importance to the observer (Mahony and Ullman, 1988). According to this theory, the visual system performs an initial low-frequency parsing of the environment to identify potential regions of interest, and assigns to each region a weight according to its saliency. Corners, high luminance, and bright colors, for example, would be assigned a high salient weight. This information is recorded in a map of the environment, which is a record of the weight of each region. The map is dynamic in the sense that recent targets are depressed as the individual moves around in the environment to prevent locking, and the weights of the regions are adjusted as their relative saliency changes.
2.3.2 Feature Integration Theory of Attention

What is the purpose of focused attention? According to feature integration theory, elementary features in the environment such as color and shape are processed before objects that require a conjunction of several features, such as a blue box or a gray kitten. Focused attention is necessary to conjoin the separate features, which then enables proper identification of the object (Treisman & Gelade, 1980).

The studies Treisman and Gelade conducted were based on the experimental paradigm known as visual search. In this paradigm, the amount of time it takes to complete a search is plotted as a function of the number of items to be searched. A flat response indicates a fast, parallel process, whereas a linear response indicates a slower sequential process. Since eye movements are inherently sequential, a task that requires eye movements would elicit longer search times for a larger number of items and a linear response.

The experiments were designed to distinguish between features that are elementary, or integral, and features that are separable and require focused attention for integration. They hypothesized that an integral feature would elicit a flat search response and would exhibit "pop-out" in a field of distractors, whereas an object with separable features would require a linear search response. Their results showed this to be the case when the elementary features were chosen to be colors or shapes (for example "pink" in a field of "brown" and "purple" distractors, or "O" in a field of "N" and "T" distractors) and the separable features were chosen to be a conjunction of the two elementary features (such as "pink O" in a field of "green O" and "pink N" distractors).

Both the saliency map theory and the feature integration theory describe perception as being the result of low-level and early-vision processes. Oculomotor behavior is a response to a stimulus provided by the attractor.
2.4 The World as Anchor

2.4.1 Semantic Consistency

When subjects are shown a line drawing of a natural scene that contains either a semantically consistent object (a tea kettle in a kitchen) or a semantically inconsistent object (a microscope in a kitchen), they are quicker to locate the consistent object, when asked to search for it, than they are the inconsistent object (Henderson, et al, 1999). Moreover, the initial saccade is equally likely to be to the consistent object as it is to be to the inconsistent object. Since the inconsistent object would seem to have a higher salience than the consistent object, the saliency map framework for early visual processing is either wrong or incomplete. A determination of semantic consistency necessarily takes into account the relevancy of a particular object in its surroundings, and this is not considered as part of the saliency map model.

2.4.2 Change Blindness

Change blindness refers to the phenomenon that occurs when large-scale changes in the visual scene go undetected by the observer as the result of a blink, a saccade, or some other visual transient. This has been explained by suggesting that attention is being prevented from being focused on the change because of the distraction caused by the visual transient. In other words, the change blindness could be due to a masking, or resetting, of the internal representation of the world (Rensink, O'Regan, and Clark, 1995). It has also recently been found that small random changes in the scene, such as that due to a mud-splash on a car windshield, can also result in change blindness (O'Regan, Rensink, and Clark, 1999). Not only are mental images unreliable, but the internal representation is quite sparse and contains only the information about the environment that is of central interest. This suggests that while feature integration and saliency mapping may be the mechanism for
encoding visual primitives and binding them together, cognition dictates what is actually preserved in and retrieved from memory. It may be that it is a more efficient strategy to use the world as an external memory source, and only encode the information that currently has meaning.

2.4.3 Exocentric Reference Frames

The notion of "world-as-anchor" can best be summed up by saying that we are perceptually predisposed to seeing the world around us as stable, despite large changes in eye and body position that displace the image on the retina significantly.

When a small afterimage is viewed in darkness except for the glow of a small, stationary reference light, and the eye moves, the afterimage appears to move relative to the reference light. When the afterimage is large (complex scene), it appears to remain stationary when the eye is moved, and the reference light instead appears to move, even though the subject knows the reference light is actually stationary (Pelz and Hayhoe, 1995). When the subjects were instructed to inspect the afterimage and made large saccades (up to 5°), the large afterimage still did not appear to move. This was explained by suggesting that whole-scene afterimages carry more perceptual "weight" than do small, isolated patches of light in a darkened room. The large afterimage creates an external reference frame, or anchor, that allows for visual stability and constancy of visual direction.

2.4.4 Position Constancy During Passive Movement

Position constancy refers to the perception that the environment does not appear to move when the eyes, head, or body moves, even though the image on the retina is displaced. Irvin Rock (1967) found that external frames of reference are used to maintain position constancy for subjects that are passively moving through their environment.
He seated subjects, blindfolded, in a small motorized wagon and started the wagon moving. He disguised the effect of the acceleration of the wagon by telling the subjects to expect a small amount of jostling of the equipment while the experiment was being set up. He then sent the wagon rolling along a darkened hallway and removed the blindfold. The only objects visible to the subject were small, luminous circles, placed along the walls of the hallway so that only one circle was visible to the subject at a time. The subjects were asked to report what they were experiencing. Seventeen of the 20 subjects reported that they were stationary and the circles were moving past them. He then repeated the experiment with different subjects, changing the luminous circles to luminous vertical lines. This time the subjects were able to see all of the lines, which filled the visual field. Twelve of the 20 subjects experienced the lines as stationary and themselves as moving. Rock explained this by saying that the lines provided a frame of reference for the subjects that enabled the correct perception. The results from this study showed that for subjects who are passively moving through their environment, position constancy can be maintained by having an external frame of reference.

2.5 Perceiving the Direction of Heading During Motion

Position constancy is not the only issue relating to the perception of a stable world in the presence of image motion on the retina. Perceiving one's direction of heading while making whole body movements, head movements, and eye movements is critical for survival in the world, and is a natural ability for humans.

2.5.1 Retinal vs. Extra-retinal Information

In the 1980's and early 1990's researchers considered the question of how people maintain their heading direction while making head and eye movements as they move about.
In this case the flow field, which results from the changing structure of the ambient optic array as the observer moves around, must be decomposed into both a translational and a rotational component. It was assumed that the rotational component due to an eye or head movement was effectively canceled out prior to the determination of heading. Several hypotheses have been proposed to explain how the rotational component could be canceled out.

The retinal image theory claims that there is enough information in the retinal image alone to accurately predict direction in the presence of head or eye movements (Warren and Hannon, 1988). The extra-retinal theory claims that proprioceptive information, and possibly an efference copy of the eye command, is necessary to make an accurate determination of direction (Royden, Banks, and Crowell, 1992).

Both theories base their claims on the results of an experimental setup that requires test subjects to view a random-dot display of simulated motion on a video screen. There are two parts to the experiment. For the first part, subjects initially fixate a central target, then pursue the target as it moves laterally across the screen. For the second part, the subjects again fixate a central target, and continue to fixate the target as the display changes to simulate a lateral eye movement. The resulting flow on the retina should be the same for both cases.

![Flow field on screen](image)

![Flow field on retina](image)

**Figure 2-5.** Experimental conditions for sufficiency of retinal information. In a) the subject was instructed to fixate the cross then make an eye movement in the direction of the arrow. In c) the subject was instructed to fixate the cross while the display simulated motion to the right. In b) and d) the resulting flow on the retina is shown. Note they are the same. (Palmer, 1999)
The major difference between the two studies was that the retinal image proponents used slow speeds for the real and the simulated eye movements (0.2° - 1.2° per second), whereas the extra-retinal proponents used faster speeds (1° to 5° per second). At the end of each 1250 msec trial, the subjects were asked to state their perceived direction of heading. Warren & Hannon instructed the subjects to indicate their perceived direction of heading by having them state whether they felt as if they were headed to the right or to the left of a vertical target line placed on the horizon of the last frame of the display after the motion had stopped. Royden et al. had the subjects state which one of the seven equally spaced (4° apart) targets was closest to the perceived direction of heading after the motion had stopped.

The retinal image proponents found no difference in perceived direction for the real or simulated eye movement. This suggests that all of the information that is required to perceive direction is present in the retinal image. Interestingly, the extra-retinal proponents discovered that there was a significant difference in perceived direction for the real and simulated eye movements. They found that the subjects could not tell in which direction they were headed without making a real eye movement. When the eye movement was simulated on the display screen (by sweeping the dot pattern laterally across the screen), they felt as if they were moving along a curvilinear path, rather than straight ahead. This is evidence that extra-retinal information is necessary for determining heading. It appears that the speed of the eye movements might be a reason for the discrepancy between the results. It is possible that at slow speeds heading information can be recovered.
from the retinal flow pattern without any extra-retinal input, whereas faster speeds require the extra information.

2.5.2 Differential Motion Parallax

In 1992 James Cutting disputed the hypothesis that moving observers decompose retinal flow into translational and rotational components. He maintained that retinal flow in its entirety is sufficient for this, in the form of differential motion parallax. He argued that the earlier studies did not include the components of bounce and sway that people experience when they move at a pedestrian speed. He reasoned that if these components are included in the experimental conditions, subjects would find it much more difficult to determine their heading direction because the additional decomposition due to these components would complicate the process of perception. He found that subjects were equally able to determine their heading direction with or without the added components of bounce and sway, and concluded that individuals use retinal information directly in the form of differential motion parallax.

Neither retinal decomposition nor differential motion parallax considers the possibility that subjects may be using the environment as an external frame of reference, in much the same sense as was shown for position constancy and exocentric frames of reference for afterimages.

2.6 The Effects of Freeing the Head

The studies conducted on heading perception discussed in the previous section were conducted in the confines of a laboratory setting. The subjects’ eye movements were monitored with a head-mounted limbus eye-tracker as they watched a simulated display of motion (dots or schematic trees floating across the screen) on a small computer monitor.
Since the human visual system did not evolve motion perception capabilities in this type of setting, it seems reasonable to conclude that the results may differ when natural movement through the real world is considered.

Traditionally, most studies of oculomotor behavior have relied upon eye-movement recording devices that required the head to be immobilized during the experiment. Usually a chin-rest or bite board was used. The reason for using a head-restraining mechanism is because in order for an accurate measurement of maintained fixation to be made, the device must be able to distinguish between motion of the eye in the head, and motion of the tracker with respect to the head. If the tracker moves while the subject is maintaining fixation, an eye movement will appear to have been made, when in reality the eye may not have moved at all. It is necessary to keep the head secured to eliminate any motion of the tracker relative to the head in order to determine when the eye is rotating.

Early eye movement monitors such as the contact-lens optical lever, the magnetic field sensor coil, and the SRI Dual Purkinje Image Tracker required immobilization of the head (Kowler, 1995). Researchers generally assumed that fixations made with the head immobilized would be the same in terms of stability as fixations made when the head was free to move but did not move. It has subsequently been discovered that this assumption is incorrect (Skavenski, et al, 1979). When subjects maintained fixation on a distant target, retinal image stabilization decreased when the head was not supported, as shown below in Figure 2-6 for two subjects.

When the head is free to move but does not, image motion on the retina can be as much as 2 or 3 degrees per second. Visual perception is insensitive to this type of motion, and it has been suggested that vision can actually be impaired when the head is not free to move (Kowler, 1995). It could be that the visual system has evolved a tolerance for retinal
image motion, to make the task of perception less taxing, much in the same way that saccade target position during reading can be very imprecise within a word, yet comprehension does not suffer. Developers of robust robotic vision might well consider the wide tolerances of human vision to be a model for systems that require the synthesis of large amounts of data for the performance of complex tasks.

Subject A

Subject B

Figure 2-6. The effect of freeing the head on fixation stability. The vertical lines represent 1 second intervals. The vertical bar on the right of each graph indicates 1 degree of rotation. (Skavenski, et al., 1979)

Another effect of freeing the head is that gaze-shifts (saccades) are faster and more accurate when the head is free to move (Collewijn, et al, 1992). Figure 2-7 shows a comparison between the position and velocity of a single gaze-shift with the head free to move and when the head is immobilized on a bite bar. This suggests that performance is more effective under natural conditions where the head is free to move.
Figure 2-7. The effects of freeing the head on saccades. The top trace (position) shows greater accuracy when the head is free (solid line) than when it is stabilized (dotted line). The bottom trace (velocity) shows that saccades are faster when the head is free to move. Note that the higher velocity is not due to head motion because the head moves after the gaze shift occurs. (Collewijn, et. al, 1992).

Finally, it has been discovered that the velocity of vergence eye movements under natural conditions can reach up to 200° per second, which is much faster than previously thought possible (Steinman, Kowler, and Collewijn, 1990). This is evidence that vergence may actually belong within the saccadic classification of eye movements, rather than smooth, and gives further reason to believe that the original classification scheme is arbitrary.

2.7 Natural Tasks

The effects of freeing the head showed that the environmental context within which subjects perform oculomotor tasks is important. Of greater importance is the nature of the task to be performed; natural tasks provide a context similar to that of a natural environment.
2.7.1 Memory Representation of a Simple Natural Task - Blocks Copying

Subjects were given the task of copying a pattern of colored blocks from a model to a workspace. It was found that they did not rely much on memory of the model, but rather frequently checked the model to acquire the information at the time that it was needed (Ballard, Hayhoe, and Pelz, 1995).

The subject was instructed to "copy the model pattern as quickly and accurately as possible." The display consisted of a model area, a resource area from which the subject could select blocks as needed, and a work area where the assembly of the copy took place. For one experiment, the display was a computer screen and the subjects used a mouse to move the blocks from the resource area to the work area. For this experiment the subject's head was held fixed on a bite bar. Another experiment allowed the subject's head to move, and the display was a set of Duplo blocks that the subject physically manipulated to copy the model.

![Figure 2-8 Blocks copying task. This is the display that was shown on the computer screen. The display subtended an area of 17 x 13° visual angle. A trace of the eye movement and of the hand movement is also shown. (Hayhoe, Ballard, and Pelz, 1994).](image)

On average, the number of fixations in the model area was 1.6 per colored block, and was somewhat lower (1 per block) when the experiment was repeated with single color
blocks. In general, it was found that subjects use one of four strategies for copying the blocks: MPMD (model-pickup-model-drop), MPD (model-pickup-drop), PMD (pickup-model-drop), or PD (pickup-drop). Model refers to a visual reference to the model area of the display, pickup refers to a visual reference to the resource area, and drop refers to a visual reference to the workspace area of the display.

![Diagram of eye movement strategies](image)

**Figure 2-9.** Eye movement strategies used for blocks copying task. Relative frequencies of each strategy category are for a sample containing approximately 50 block moves for each of seven subjects. (Hayhoe, Ballard, and Pelz, 1994).

From Figure 2-9, it is seen that the MPMD strategy was the most frequently used. This strategy represents a relatively "memoryless" sequence, in the sense that a separate fixation is used to identify the color of the block and the location of the block; they are not remembered together, as a unit. On the surface, this strategy seems to lend credence to
Treisman's feature integration theory of separable and integral features because color and location are identified as elementary features during execution of the task. However, in this case a fixation does not seem able to fully bind the features together because separate fixations are usually needed to capture each of the elementary features. Incomplete feature integration is occurring during fixations for this task.

This strategy is also an example of a "deictic" strategy, one in which a reference or pointer to the information is maintained, rather than the entire structure. A complex internal representation of the scene can be avoided by employing this type of strategy - yet another example of how vision and cognition work in synchrony to make the task of perception tractable.

### 2.7.2 Sequential Looking Task - Tapping vs. Looking Only

Gaze-shift dynamics have been shown to be controlled, in least in part, by the demands of the visual task being performed. This conclusion was reached by studying the oculomotor behavior of subjects while they were engaged in two different sequential looking tasks (Epelboim, et al, 1997).

The experiment consisted of having a seated subject locate a sequence of small, colored lights (5mm diameter LEDs) that were mounted on a worktable in front of him or her. A target sequence consisted of either 2, 4, or 6 lights. The order of the sequence was known by the subject before each trial began, but the subject's eyes were closed while the lights were positioned on the table so he or she would not know the specific location of each light in the sequence. Both the head and torso were free to move during the experiment.

For the first condition the subject was asked to look at each target light as quickly as possible in the correct sequence. For the second condition the subject was asked to tap each
target as quickly as possible in the correct sequence. The two conditions were performed in separate sessions on separate days.

The results, as indicated in Figure 2-10, show that both gaze-shift and eye-in-head peak velocities are greater for tapping than they are for looking only. Gaze shift refers to the rotation and translation of the head combined with rotation of the eye, whereas eye-in-head refers to the rotation of the eye relative to the head.

Figure 2-10. Tapping vs. Looking-only. A) Peak velocity of gaze-shift. B) Peak velocity of eye-in-head. C) Fixation duration as a function of gaze-shift amplitude for tapping (dark circles) and looking-only (light circles). Epelboin, 1997.

Why are saccades faster and fixations shorter for tapping than they are for just looking? Epelboim et al (1997) offers two possible explanations. One is that a reduction in the gain of the VOR during tapping is responsible for the faster gaze-shifts. If VOR is reduced, then movement of the head is not completely compensated for by a counter-
rotation of the eye and both gaze and eye-in-head velocities would be higher. Thus, the type of task to be performed can enhance oculomotor functioning.

Another explanation is that the oculomotor system programs eye and head movements together, and the synchronized movement is further supported by additional movements of the arms and torso to provide a highly efficient and effective means of performing complex tasks. In this sense, tapping can be thought of as streamlined motion with the arms, head, torso and eyes reaching for the target in tandem. Support for this idea is evidenced from the observation that during looking-only the arms and torso did not move at all, and head movements were smaller and slower. It is also consistent with the findings from Collewijn et al (1992) who showed that the inhibition of head movements resulted in slower peak velocities and less accurate saccade landings.

During tapping, gaze was never stationary – retinal image velocities reached as high as 4 degrees per second during fixations. However during looking-only gaze was much more stable, with retinal image velocities almost never exceeding 1.5 degrees per second. This is consistent with Skavenski et al (1979) who suggested that retinal image motion is adjusted so as to be optimal for natural body motion, and for the requirements of the task at hand. It may be that the higher tolerance for image motion during tapping is sufficient for a task that does not require high visual acuity.

2.7.3 Visual Memory in Problem Solving – Geometry

Memory representation in a natural task has already been discussed for a block-copying task. It was found that subjects chose to use very limited working memory for this type of task, and tended to make multiple fixations on the same target to acquire information when it was needed. In a separate study, subjects were found to have a limited memory
capability of approximately five items for solving geometry problems (Epelboim and Suppes, 1999).

Three subjects were asked to solve a set of simple geometry problems. Two of the subjects were considered to be experts, the other was considered a novice. Their eye movements were monitored while they solved the problems, and they were asked to verbalize the procedure they undertook. The verbal protocol showed that eye movement data provides additional information into the nature of certain intellectual tasks, such as mathematical reasoning. This information is not apparent from the verbal description. For example, the subjects did not always mention certain parts of the diagram they had fixated, and they scanned some parts several times, not mentioning why they were rescanning.

Eye movements and the cognitive processes that underlie them are closely bound together. This study showed that it is possible to infer cognitive processes from eye movement patterns, and that these processes are not always amenable to conscious verbalization. This study also showed that the size of working memory is highly dependent upon the type of task being performed.

2.7.4 Eye Movements and Work Load During Driving

Copying patterns of colored blocks, tapping sequences of colored lights, and solving geometry problems can surely be considered natural tasks when compared to watching isolated spots of light in the darkness, or simulating motion with schematic trees on an oscilloscope screen. Another natural task of higher relative complexity than those already mentioned is that of driving a motor vehicle through town.

Unema and Rötting (1988) studied the difference in eye movements and mental workload for experienced and inexperienced drivers. Data was collected for 20 city bus
drivers in Berlin for a course that was divided into segments of varying difficulty. Of the 20 subjects, 5 were instructors and 15 were trainees.

It was found that the instructors had longer average fixation than did the trainees, as shown in Figure 2-11.

![Figure 2-11. Fixation durations for expert and novice drivers, for several conditions. Unema and Rötting, 1988.](image)

The Novice 1 group refers to the trainees before they had completed the training program, and the Novices 2 group refers to the same trainees after they had completed the training program. Oddly, the trainees had shorter fixations after completing training, which was less similar to the data found for the experts. This anomaly was explained by noting that the data for this condition was collected just prior to the final trainee examination, and the trainees may have been overly nervous during this condition which might have affected their performance.

Another interesting finding was that fixation durations decreased with increasing situation complexity. Fixation durations for the easiest condition (highway driving) averaged 279 msec over all subjects (both experts and novices), and averaged 213 msec for the most difficult condition (roundabout). This is shown in Figure 2-12.
To summarize their findings, Unema and Rötting showed that fixation durations could be used as a metric to distinguish between different levels of perceived cognitive complexity. They presented their findings in terms of mental workload; a high mental workload is characterized by short fixations, and in general is preceded by a long saccade. Again, it has been shown that the oculomotor system is programmed in a way that is dependent upon high-level cognitive strategies.

2.7.5 The Direction of Gaze During Driving

Recordings of subjects' direction of gaze (eye-in-head coordinates added to head-in-car coordinates) and steering wheel angle while driving along a winding road showed that subjects frequently look at the point of tangency on the inside of the bend (Land and Lee, 1994). The point of tangency is defined to be the point on the curve that is changing direction as the driver approaches it, and is relative to the position of the driver.
The drivers tended to look repeatedly at this point while they were negotiating a curve (80% of fixations) and sought this point out 1–2 seconds before entering into the curve. They also returned to this point repeatedly after entering the curve, as a point of reference. Moreover, the drivers consistently turned their steering wheel toward the direction of gaze approximately 0.75 seconds after establishing the gaze point, and maintained fixation there for approximately 3 seconds into the bend.

Land and Lee (1994) suggested that this behavior is part of an overall steering strategy that drivers employ that enable them to predict the curvature of the road up ahead. It is also another example of a deictic, or "do-it-where-you-look", strategy because the eye moves so as to be able to capture the most relevant information in the visual field for performing the current task. This type of strategy greatly simplifies the task to be performed, because there is no need to keep track of the absolute position of an object being manipulated. The relative position, centered at the point of fixation, is sufficient for the successful completion of the task.

2.7.6 Eye Movements While Making Tea

The studies on eye movements while driving (Land and Lee, 1994) lead one to consider how the control of other everyday activities is monitored by eye movements. Are the eyes essentially passive "cameras", capturing images of the world onto the retina and passing along this information to the brain via the optic nerve, or is there a tighter coupling between plans, actions, and oculomotor behavior to enable the efficient execution of complex tasks?

Subjects' eye movements were monitored while they performed the everyday over-learned (for them) task of making a pot of tea and pouring themselves a cup (Land, Mennie
and Rusted, 1999). The setting was an ordinary yet unfamiliar kitchen where the subject had to search for the necessary items to complete the task.

Eye movements were recorded using a portable eye-tracking device that consisted of a videocamera mounted onto a band from a construction helmet. The band was secured to the subject’s head, but the head was free to move naturally. The upper two-thirds of the camera’s field of view imaged the kitchen scene, and the lower third imaged the subject’s left eye. The scene and the eye were recorded together with a videorecorder that was placed inside of a backpack. The subject wore the backpack while they performed the task. At a later time, the recorded image of the eye was calibrated to correspond to eye-in-head rotation, and a marker was superimposed over the scene to indicate point of gaze.

Land et al (1999) described the execution of the task of making tea as being part of a hierarchy of goals. The overall goal is that of the task itself – making tea. The finest level of the hierarchy corresponds to the sequence of fixations necessary to support the higher levels of sub-goals. When the patterns of fixations were considered in the context of supporting sub-goals, it was found that nearly all of the fixations (over 95%) were directly related to an object manipulation or eventual manipulation.

When actions are under visual control, eye movements serve the purpose of not only monitoring actions as they unfold, but also serve as an explicit manifestation of the cognitive script employed to dispatch a high-level strategy. For example, this experiment found that some actions are embedded in others, such as removing the lid from the kettle as the kettle was carried to the sink. Fixations were made on both the lid and the sink during that time, showing that the action of removing the lid was embedded in the larger action of getting the kettle to the sink. Both actions were under visual control, not just the main action.
Also, time-sharing of tasks was evident upon inspection of the fixation patterns. When one hand was swirling the teapot, the other hand was replacing the top on the milk. Fixations were alternating between the two sub-tasks during this time. Ordinarily we are well aware of our ability to time-share between sub-tasks in this way, however it was not obvious previously that this cognitive strategy was evident so explicitly in eye movement patterns.

The temporal relationship between vision and action was also documented. On average, the trunk of the body preceded a saccade to the object to be manipulated by 0.63 seconds, and the saccade preceded the first sign of physical manipulation by 0.57 seconds. The majority of fixation durations were between 0.2 and 0.5 seconds, however there were some very long fixations associated with waiting for something to happen. Saccade sizes peaked at around 5°, with many very large saccades, bringing the average saccade size up to 20.2°. It was noted by Land et al (1999) that saccades this long are not typical for other tasks such as reading text, where typical saccade sizes range from 1 to 2°. The longer saccades were associated with the searching sub-tasks.

The searching sub-task also showed several fixations on objects that were not required for the current sub-task, but that were used at a later time. For example, it was reported that the box of sweeteners was located twice during an undirected search, but was not used until 68 seconds later. The authors hypothesized that the spatial coordinates locating the sweetener were being preserved in memory to enhance a strategy for accurate retrieval at a later time. This strategy was apparently also used to replace items after they were used — "interestingly, when the sweeteners were put back on the shelf, after an interval of only 5s, gaze preceded their return to the exact point on the shelf from which they had come" (Land, et al, 1999).
Overall, it was found that the eyes monitored and also guided virtually every action that was necessary to complete the task of making tea. The salient properties of objects were not as important for fixation as was their relevance to the task at hand. This is a consequence of the goal-driven behavior of the subjects; as Yarbus (1967) noted earlier, patterns of fixation scan-paths are highly dependent upon the goal of the observer. Salience is an important indicator for determining which object is fixated next for free-viewing but not for planned behavior that requires a strategy for execution.

Most human behaviors, even over-learned ones such as making a pot of tea, are goal-oriented and require a strategy. Even when humans are engaged in passive activities such as watching television, they have expectations and thoughts about the characters and plots that influence their visual patterns, and thus their perception of the events. Therefore, the central idea surrounding this thesis is that it is more appropriate to study visual perception from a top-down perspective, that is, to first consider the cognitive state of the individual and the context of the environment, and then consider the low-level implementation details to be a support mechanism for the higher processes. This approach has several significant advantages — one is that the "problem" of vision is greatly constrained; we only need to deploy neural resources as they are needed, and in a way that is "good enough" to get the job done. The second is that this approach provides the rationale for discarding the linear systems approach of understanding vision as a sum of simple processes. On the contrary, the low-level processes should be thought of as a consequence of, rather than the contributors to, high-level perception.

2.7.7 Portable Eye-Tracking and Natural Tasks

The idea of studying eye movements in the context of complex, unrestrained behavior in a natural environment was extended by Pelz et al (2000). Four complex tasks were
studied using portable eye-tracking equipment similar to the device used by Land et al (1999) for studying tea-making. The tasks showed increasing levels of overall complexity, from the relatively simple task of image-quality judgements, to the more cognitively complex tasks of searching a road map, building a model of a toy rocket, and washing one's hands in a restroom.

Each of the tasks showed distinctive patterns of eye movement behavior and associated metrics for fixation durations and saccade sizes. All of the tasks showed several instances of extremely short fixations (~33 msec) which had only been found in previous studies designed to elicit a specific response, such as express saccades to an anticipated target, or as part of a two-part saccadic sequence (Epelboim, et al, 1997).

Some of the tasks were found to be decomposable into distinct sub-tasks, for which oculomotor behavior seemed to be tailored specifically for optimal performance of the particular sub-task. For example, building a model of a rocket could be parsed into actions involving either reading text, searching for a model part, or manipulating the parts to build the rocket. Searching for a part elicited the shortest fixations and the longest saccade lengths, whereas manipulations elicited the longest fixations.

The handwashing task showed several instances of "planning," or "look-ahead" eye movements, where the subject located an object much earlier than was required for manipulation of the object. This strategy seemed to optimize the performance of the task in much the same way as noted earlier by Land et al (1999).

The study of eye movements in natural environments, with unrestrained subjects, is clearly in its infancy. Much work needs to be done to enhance our understanding of how visual perception, oculomotor behavior, and the planning of goal-oriented tasks interact with one another to produce both spatial and temporal coherency with limited neural hardware.
A description of how each component works in isolation is insufficient for understanding the complete organism. A task-oriented approach, as suggested earlier, will shed more light on how the visual system is capable of providing relevant information to the individual in an efficient and timely manner. The lesson provided by the background material for this thesis is that eye movements are task dependent and can reveal cognitive strategies. Visual capabilities, rather than being general and insensitive to context, are in essence specific and tailored to the requirements of cognition.
Chapter 3

3. Approach

The primary focus of this research project is to investigate the oculomotor behavior of subjects while they are performing everyday tasks in a natural environment. The investigation consisted of collecting eye movement data in the form of a series of eye-tracking experiments designed to elicit natural responses in a non-laboratory environment. Very little information exists about this topic, primarily because the technology that is required to capture this type of eye movement data has not been available to researchers until recently.

In order to carry out this research effort, and to enable other ongoing research projects on this topic, the Visual Perception Laboratory at the Carlson Center for Imaging Science of the Rochester Institute of Technology has developed a portable, wearable eye-tracking system. This system is based on, and is an extension of, the Applied Science Laboratory model 5000 eye-tracking system and model 501 head-mounted optics. All of the equipment needed to carry out the experimentation and analyze the data is located in the Visual Perception Laboratory at RIT.
3.1 History of Eye-Tracking Methods

According to Arrington (2000), eye-tracking methods can be classified into broad categories: electrical methods and optical methods.

3.1.1 Electrical Methods

Electro-oculography (EOG) is the measurement of the electrical potential between the front and the back of the eyeball. When the eye rotates, the orientation of this potential changes, and this change can be used to determine when the point of gaze changes and also the amplitude and direction of the change. This technique is not very precise, however, as there is a significant amount of crosstalk between the horizontal and vertical components, and it also is not very accurate due to a substantial drift of the signal over time. It is necessary to conduct experiments at the same time every day for a particular subject because the potential varies over the course of the day. It is also very uncomfortable for the subject. This technique is rarely used because of these significant drawbacks.

During the 1960's and 1970's the magnetic field induction coil method of eye-tracking came into widespread use. The subject is placed inside a room that contains large magnetic induction coils. This bathes the head in a weak alternating magnetic field. The subject's eye is anesthetized and a tight fitting corneal annulus with a coil of wire attached to it is placed on the eye. When the eye rotates, the coil moves and a voltage is induced in the wire due to the alternating magnetic field. The voltage is converted to a current, which indicates the amount the eye has moved. This method usually necessitates securing the head on a bite-board or chin rest.

The Maryland Revolving-Field Monitor (MRFM) bypasses this restriction by using a Sparker Tracking System (STS) that tracks 3-D translations of the head. In addition to the eye coil, a sparker coil is secured on the head and generates acoustic signals whose arrival
time to the STS is detected. Both induction coil methods are highly accurate and have very high temporal resolution. The single coil method is also frequently used for animal experiments, where the wire coil is permanently implanted in the animal's eye for even greater accuracy.

3.1.2 Optical Methods

The contact-lens optical lever is an instrument that was used in the early studies of maintained fixation. A contact lens with an attached mirror is fitted over the white sclera of the eye, and a narrow beam of light is shined onto the eye. It requires securing the head so as not to confound eye rotations with head translations for an accurate indication of gaze-point. When the eye moves, the light that is reflected off of the mirror is shifted slightly, and this shift corresponds to twice the amount the eye has moved.

Another optical method of tracking the eye involves measuring how much the cornea bulges as the eye rotates. An array of infrared detectors is placed around the eye, and when the cornea bulges, they sense the variation in the light that is reflected back from a source shined on the eye. This method is also sensitive to head movement, so the head must be secured.

Other methods of detecting eye movements rely on the reflective properties of the eye itself. The limbus tracker works by measuring the difference between the amount of light that is reflected from the cornea and from the sclera. The junction between these two areas is called the limbus, and is used as the boundary for measuring the contrast between the two areas. To detect horizontal movement, two photodetectors are placed on the right and left sides of the eyeball to detect the contrast change when the eye moves. To detect vertical movement, the photodetectors are placed on the lower eyelid. There can be some crosstalk
between the horizontal and vertical components, but overall the method is advantageous because it is a fast, analog method of eye-tracking and exhibits high temporal frequency.

Other reflective techniques are based on the Purkinje images, which are reflections from the various surfaces of the eye. The first through fourth Purkinje images are reflections from, respectively, the front surface of the cornea, the rear surface of the cornea, the front surface of the lens, and the rear surface of the lens. The higher order Purkinje images may be very faint and difficult to detect.

The first Purkinje image, commonly referred to as glint, can be used alone to detect the direction of gaze, but this method is highly sensitive to head movement. To circumvent the necessity of having the head securely restrained, the first Purkinje image can be used in combination with other reflections from the eye to determine the point of gaze. These methods rely on the absolute difference between the two points, and are known as vector difference methods. The Dual Purkinje Image Tracker uses both the first and fourth images to detect eye rotation with respect to the head, but since the fourth image is very faint, firm head support is still necessary to achieve an accurate measurement. This method also suffers from a limited range of eye movement detection.

A variation of this vector difference method is to use the first Purkinje image as the first point, and the center of the pupil as the second point. This method works without having the head secured because the distance between these two points remains constant whenever the head moves but the eye does not. Thus, detecting a change in this distance indicates an eye movement with respect to the head. For example, when the head moves but the eye does not, both the pupil center and the corneal reflection (first Purkinje image) move by the same amount for a fixed light source. When only the eye moves, the pupil center moves significantly but the corneal reflection moves only slightly (due to the
curvature of the cornea). Thus, if the difference between the pupil center and the corneal reflection is represented by a vector, a change in the magnitude of the vector corresponds to the amplitude of the eye movement with respect to the head, and a change in the angle of the vector corresponds to a change in the direction of gaze. Since a head movement alone is represented by a vector of constant magnitude, this constant can be factored out to discount the head movement.

There are several disadvantages to using the vector difference method. One is that there are now two sources of noise for the gaze computation, one from each reflection. Related to this is the lowering of the signal-to-noise ratio. This comes about as a result of the signal from the vector difference being slightly smaller than the separate signals coming from the two points combined (Arrington, 00). Another disadvantage of this method is that it is sensitive to the distance between the detector (camera) and the head. As the detector moves further away from the head, the vector distance becomes smaller for a constant gaze shift. This is not a problem if the detector is fixed relative to the head. Finally, this method is slower than the other methods because the sampling rate is effectively limited by the video capture capabilities of the detector, which is usually a 60 Hz field rate.

3.2 The VPL Portable, Wearable Eye-Tracker

The Visual Perception Laboratory (VPL) at RIT has developed a head-mounted goggle system that is used for eye-tracking outside of the confines of the laboratory setting. It uses the vector difference method of eye-tracking as described above, the specifics of which will be explained in more detail in the following sections. The development of the goggle system was based on the Applied Science Laboratories (ASL) model 501 head-mounted optics module, and uses the ASL E5000 eye-tracking software to compute the line of gaze. The
software also implements a user interface program to enable system calibration and to allow for communication with the software.

The entire hardware system used for portable eye-tracking consists of the custom goggles headgear, a portable control unit with power supply, two Hi-8mm camcorders with batteries and videotapes, a Pentium PC notebook computer, an optional microphone for recording audio, cables for connecting the control unit to the eye-tracker, the notebook, and to the two camcorders, a nylon backpack to house the control unit, camcorders, and power supply, and an electronic flash unit for synchronizing the scene and eye videorecordings.

3.2.1 The Custom Goggles Headgear

The wearable headgear consists of an optics module with an eye camera, a small mirror, a scene camera, and a small LASER mounted on a lightweight plastic racquetball goggle system meant to be worn on the subject’s head, as shown in Figure 3-1. To ensure proper fit, and to eliminate unnecessary jostling of the components, a strap at the back of the goggles can be adjusted.

Figure 3-1. Portable, wearable eye-tracking headgear
Optics Module and Mirror

Figure 3-2 shows the optical path for the components of the portable eye-tracking headgear. The optics module (HMO) consists of a CMOS eye camera, a near-infrared LED (IRED), a beamsplitter, lenses, and a prism to direct an image of the illuminated eye into the eye camera. The purpose of the optics module is to focus an image of the left eye onto the solid-state CMOS sensor of the eye camera. Figure 3-3 shows a top view of the goggles system.

The IRED illuminates the eye. The eye camera is able to capture the partially collimated beam of light that is reflected off the back of the retina because the optical axis of the eye camera is coaxial with the illuminating beam. This allows the image of the pupil to be back-lit bright, rather than dark, and affords an easier discrimination of the pupil from the iris and other dark parts of the eye.

A mirror is contained inside a prism housing to direct the optical path through the eye camera lens. The illuminating beam is aligned with the camera optical axis by means of adjustment screws that rotate a partially silvered beamsplitting mirror. The IRED housing connects to the ASL control unit by means of two connecting leads.
External to the optics module is a small IR reflecting mirror that is mounted on the inside of the goggles near the left eye. This mirror directs the illumination to the eye and reflects an image of the eye back to the eye camera. The area behind the retina, called the pigment epithelium, contains pigmented cells that are highly reflective in the red and infrared region. When light of the proper wavelength range strikes the retina on axis, this light is retroreflected back out of the eye along the same path it entered. The mirror then directs this retroreflected light to the eye camera inside of the optics module. The inner surface of the mirror is coated with a material that is reflective in the near IR, and transmissive to visible. It does not interfere with the subject's view of the environment.

**The Eye Camera**

The eye camera sensor consists of a solid state CMOS chip and first stage electronics for a 60 Hz monochrome video format. The sampling rate is effectively reduced to 30 Hz due to the fact that two fields are averaged for each frame. Thus, a sample is effectively recorded every 33 msec. Fixation durations as low as 33 msec are resolvable in freeze-frame mode if the fields surrounding the frame in question are looked at individually (not averaged), and they show evidence of an eye movement occurring just before and just after the fixation. The need for this level of temporal resolution is not usually necessary since fixation durations are typically not less than 100 msec (Carpenter, 1988).

A Kodak Wratten 87 filter is placed in front of the eye camera to isolate only the short wavelength infrared radiation. Camera focus is achieved by rotating a telescoping focusing tube. The tension on the tube is adjusted with a focus adjustment set screw. Cables from the camera sensor head connect directly to the ASL control unit.
**Scene Camera**

The scene camera focuses an image of the scene being viewed by the subject onto a color CMOS sensor, and creates a frame of reference for measurements of eye line of gaze. It is mounted on the goggles just above the left eye of the subject and points in the same direction that the subject is looking. This set-up creates minimal parallax error in the scene image.

The scene image is in color, and covers a field of view of 75° horizontally and 53° vertically for the camera lens with a focal length of 6mm. The camera position is fixed on the goggles, and is not adjustable. The sensor chip in the camera connects to the ASL control unit. Both the scene camera and the eye camera are powered from a single 9 volt rechargeable lithium-ion battery attached to the goggles. The original ASL head mounted optics unit uses CCD sensors for both the eye and scene cameras. The CCD sensors have a higher resolution than the CMOS sensors, however cameras with CMOS sensors are lighter and have a lower power consumption than cameras with CCD sensors. These are important considerations for portability requirements.

**LASER**

A small, visible wavelength LASER is mounted on the goggles near the right ear. The LASER is used to project an image from a 2-dimensional diffraction grating onto the subject’s field of view for system calibration purposes. Since both the LASER and the scene camera are fixed relative to the head, any small movement of the head during calibration will not be visible to the scene camera, and will not affect the software capture of the target calibration points. This allows for a robust and highly accurate computation of line of gaze, without the need to restrain the head during calibration.
3.2.2 Other System Components

Figure 3-4 is a photograph of the remaining hardware that is used to perform the eye-tracking experiments. The ASL portable control unit, 2 camcorders (one to record the eye camera video, and one to record the scene camera video), a small LCD television monitor, and a portable power supply fit into a backpack that is worn by the subject during the experiment.

![Figure 3-4. Eye-tracking hardware, not including the goggles.](image)

![Figure 3-5. Camcorder inside backpack.](image)  
![Figure 3-6. Close up front panel of control unit.](image)
A flashlamp is used to synchronize the video eye and video scene data. A Pentium PC notebook computer runs the software that communicates with the experimenter and displays the data that is computed by the ASL control unit. After a calibration of the system for a particular subject is performed, the notebook can be disconnected from the control unit and left in the laboratory while the subject performs the experimental task.

Figure 3-7. Subject wearing eye-tracking gear, ready to perform an experiment.

The purpose of the control unit is to process the eye camera signal to find the center of the pupil and the reflection of the illuminating source off the front surface of the cornea (the first Purkinje image). It receives the video signals from both the eye and scene cameras and uses this information to compute a line of gaze and display it as a cursor superimposed on the video image of the scene. The gray-level histogram of the eye image is thresholded at two levels to distinguish the pupil area from the surrounding iris. The software then computes the pupil diameter and displays it as an outline of the pupil, and the corneal
reflection diameter as an outline of the corneal reflection. Centered crosshairs are also displayed along with the outlines – white for the pupil, and black for the corneal reflection. These displays are shown on the eye camcorder screen, the LCD television monitor screen, and the notebook computer screen. The displays are recorded onto the videotape of the eye camcorder.

Figure 3-8. Image of pupil (white outline) and corneal reflection (black outline). Centers are indicated by crosshairs.

The control unit is powered by two camcorder batteries connected in series, and will run for approximately 2 hours before recharging is necessary.

3.2.3 Theory of Operation

The line of gaze is determined by measuring the separation between the center of the pupil and the center of the corneal reflection. A change in the line of gaze is approximately proportional to a change in this separation. In other words, when the eye rotates to change the line of gaze, the separation between the pupil and the corneal reflection also changes proportionately. The exact relationship is given by:
The geometry is illustrated in Figure 3-9 below.

\[ S = K \sin \theta \]  

(1)

The computation assumes a spherically shaped cornea. \( \theta \) is the eye line of gaze angle with respect to the eye camera and the IRED light source. \( K \) is the distance between the center of the pupil and the center of curvature of the cornea; both \( K \) and \( \theta \) are used to determine \( S \), the separation between the center of the pupil and the corneal reflection. \( S \) is directly proportional to the direction of gaze. This computation is done in both the horizontal and vertical direction.
The corneal reflection is detectable up to approximately 50° of visual angle field; thus, if a subject looks away from the camera at an angle greater than 25°, the corneal reflection is no longer detectable.

### 3.2.4 Eye-Tracker Set-Up and Calibration

In order to run an eye-tracking experiment, the E5000 eye-tracking software package must be loaded into the interface PC notebook computer, and be ready to run. E5000 is the software that calculates the direction of gaze, and is also the means of communication between the experimenter and the control unit.

The user interface program runs on the computer under the DOS shell and is simple and intuitive to use. The ASL Model 501 Eye-Tracker Manual contains the complete instructions for using the interface program along with a detailed listing of all the available commands and options, therefore that information will not be repeated here. However, a brief explanation of the procedure to be followed for calibrating the system and setting up an experiment is necessary in order to fully understand the capabilities and limitations of the system.

First, the eye and scene cameras are powered up and connected to the ASL control unit, which is then connected to the PC notebook computer. The control unit is turned on and the E5000 software is loaded. The illuminator is turned on with a software switch in the user interface program, and the subject then puts the goggles on. The Velcro strap on the goggles is adjusted for maximum comfort and to assure minimal slippage. A slight adjustment of the IR mirror or the optics module may be necessary to enable the eye camera to locate the pupil, and center its image on the monitor. A thresholding procedure is then performed by the experimenter to facilitate the detection of the pupil and corneal reflection (CR) outlines.
The first step of this procedure is to set the pupil and corneal reflection discriminators using the E5000 program. The discriminators define the outlines of the pupil and corneal reflection by adjusting the thresholds for the edge detection logic. Discriminator values vary from 0 to 255. A value of 0 will detect no edges, whereas a value of 255 will detect spurious noise and interpret that as an edge.

The next step is to enter the coordinates of a calibration target test pattern into memory. This process is known as a "target sweep". A target pattern consists of a set of nine points that are projected onto the subject's field of view. The points come from a 2-dimensional laser diffraction pattern. The laser is mounted on the goggles near the subject's right ear, and is fixed with respect to the head. Since the scene camera is also fixed with respect to the head, any small movement of the head during calibration will not be visible, making the target sweep robust to small head movements. The image of the diffraction pattern is captured by the scene camera. The points are spaced approximately 15 - 20° visual angle apart horizontally and vertically, as shown below in Figure 3-10.

```
  ●  ●  ●
  ●  ●  ●
  ●  ●  ●
```

*Figure 3-10. Diffraction pattern used for calibration.*

To enter the coordinates of each point of the diffraction pattern into memory, the experimenter positions the cursor over each of the points sequentially, from left to right and top to bottom, and strikes the return key of the notebook computer at each point.
The final step of the procedure is to perform a subject calibration. Different subjects have differently shaped corneas, therefore a calibration routine is required for each subject. The calibration determines the relationship between the raw data and the line of gaze for a particular subject. The subject is instructed to fixate each point of the diffraction grating sequentially for a few seconds. At each fixation the experimenter enters the coordinates of the point into memory by striking the space bar of the notebook computer at each point. A polynomial interpolation is performed by the software for points in the field of view that do not coincide with one of the nine calibration points.

3.2.5 Eye Movement Monitoring

Once the calibration procedure is complete, the ASL unit can be disconnected from the PC notebook computer, and the eye-tracking experiment can begin. The camcorders are set to begin recording videotapes of the eye and scene image. An electronic flash unit is flashed to allow for a point of reference for synchronizing the eye videotape with the scene videotape, for later data analysis. The notebook computer is small enough and light enough to be left connected to the control unit and placed inside a pouch in the backpack, if so desired. This enables eye-tracking data to be captured in real-time and stored on the computer's hard drive for later analysis. One field of data consists of vertical and horizontal eye position coordinates, pupil diameter, and 16 bits of external data (XDAT). Event marks can also be entered from the keyboard while the subject performs the experiment.

If the subject blinks while performing the experiment, the point-of-gaze (POG) cursor that is superimposed on the scene image will temporarily freeze at the current position for a maximum of 12 video fields (200 msec at 60 Hz). If the eye is still closed after 200 msec the cursor will jump to a default position of (0,0) and will not be visible on the monitor or the videotape. Since blinks are typically shorter than 200 msec, this feature has the effect of
ignoring blinks and eliminating any data that may have been collected when the eyes are closed or the image of the eye has been lost.

When the experiment is finished, the subject can remove the goggles and backpack, and the equipment can be powered down. The videotapes are removed from the camcorders for synchronization and data analysis.

3.3 Example of Real-Time Data Capture

Figure 3-11 is an example of an eye-tracking trace made while a subject was performing the calibration procedure. The subject was instructed to look at each target point, starting at the upper left corner, for approximately 2 seconds.

![Portable EyeTracker Accuracy](image)

**Figure 3-11. Real Data Capture**

Figures 3-12 and 3-13 show traces of the vertical and horizontal eye positions for the same set of data as Figure 3-11. Figures 3-14 and 3-15 show expanded views of Figures 3-12
and 3-13. From these figures it can be ascertained that the precision of the eye-tracker is 0.5° for a camera lens with a focal length of 6mm. This is due to the fact that the video is digitized, thus it is sampled and quantified at the discrete levels shown in Figures 3-14 and 3-15.

![Portable Eyetracker Vertical Position](image1)

**Figure 3-12. Trace of vertical eye position**

![Portable Eyetracker Horizontal Position](image2)

**Figure 3-13. Trace of horizontal eye position**

![Portable Eyetracker Vertical Position](image3)

**Figure 3-14. Expanded view of vertical eye position**

![Portable Eyetracker Horizontal Position](image4)

**Figure 3-15. Expanded view of horizontal eye position**

Figures 3-16 through 3-19 show the operating noise of the eyetracker, captured at 0, 2, 4, and 8 field averaging. An ASL artificial eye was used for the target to minimize any human eye tremor that might effect eyetracker noise measurements. The headgear was mounted securely at a distance of approximately 5 inches from the artificial eye to capture the illuminating beam that was reflected off of the artificial eye.
3.4 Data Analysis

After the data has been captured on videotape, the videotape is analyzed to extract the relevant information. The first stage of analysis consists of coding the videotape to identify saccade lengths and intersaccadic intervals (fixation durations). The object of the saccade
can also be identified at this stage. The second stage consists of a statistical analysis of the extracted information so that conclusions may be drawn about the relationship between eye movements and visual perception.

3.4.1 Coding the Data

A videotape of the eye image and a videotape of the scene image are captured during the experiment. To make the data analysis easier, the eye videotape is superimposed on the scene videotape in the upper right corner of the scene, covering approximately 1/16 of the scene image. The synchronization of the two videotapes is accomplished by recording both images onto a second-generation videotape using a video mixer, and starting the recording at the flashed frames.

After the synchronization, a timestamp is superimposed onto each frame of the resulting videotape. Then the actual coding process can begin. Coding refers to the process whereby each frame is manually described in terms of a given metric by locating the position and timecode of the superimposed cursor. Table 3-1 shows an example of a segment of code. The columns (fields) of the code, from left to right, refer to the line number, a millisecond reference marker for each frame, the number of milliseconds that have elapsed on the tape since the start of the coding session, the number of milliseconds since the last frame was coded (a value of zero indicates that the frame is receiving multiple comments), the actual timecode of the frame, and a comment to describe each frame. Frames are temporally resolvable every 33 milliseconds, therefore the millisecond fields are always incremented in steps of 33.

The comment field is where a description of each frame is recorded. The comment on line 1, "start", refers to the beginning of a coding session. The comment on line 2 refers to an observation about the current video frame, and line 3 indicates the object of the
fixation. Line 4, "B_FIX", refers to the beginning of a fixation, and is always followed by "E_FIX" at the end of the fixation. Line 6 indicates the size of the saccade to the next fixated object, in degrees of visual angle. Finally, line 7 refers to the object of the next fixation, and will be followed by a "B_FIX" to indicate the start of the next fixation.

<table>
<thead>
<tr>
<th>line number</th>
<th>msec</th>
<th>since_start</th>
<th>since_last</th>
<th>timecode</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32233</td>
<td>-200</td>
<td></td>
<td>00:00:32:07</td>
<td>start</td>
</tr>
<tr>
<td>2</td>
<td>32233</td>
<td>-200</td>
<td></td>
<td>00:00:32:07</td>
<td>soap disp look ahead</td>
</tr>
<tr>
<td>3</td>
<td>32233</td>
<td>-200</td>
<td></td>
<td>00:00:32:07</td>
<td>right soap disp</td>
</tr>
<tr>
<td>4</td>
<td>32233</td>
<td>-200</td>
<td></td>
<td>00:00:32:07</td>
<td>B_FIX</td>
</tr>
<tr>
<td>5</td>
<td>32566</td>
<td>133</td>
<td></td>
<td>00:00:32:17</td>
<td>E_FIX</td>
</tr>
<tr>
<td>6</td>
<td>32633</td>
<td>200</td>
<td></td>
<td>00:00:32:19</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>32633</td>
<td>200</td>
<td></td>
<td>00:00:32:19</td>
<td>right faucet</td>
</tr>
</tbody>
</table>

Table 3-1. Example of code from eye-tracking experiment

3.4.2 Fixation Durations and Saccade Size

Fixation Duration

The duration of a fixation is defined as the time interval between the landing of one saccade and the initiation of the next. A fixation duration is given in units of milliseconds and is resolvable every 33 milliseconds. They are typically on the order of 250-350 msec, although some can be as long as several seconds depending on the type of task being performed. For example, tasks that require the manipulation of parts such as building a model of a toy rocket, or tasks that require close scrutiny of objects such as threading a needle, may have fixations as long as several seconds (Kowler, 1995).

Fixation durations are extracted from the coded data by subtracting the timecode of the videoframe labeled "B_FIX" from the timecode of the videoframe labeled "E_FIX". Since these labels are assigned by a human operator during the coding procedure, they are subjective in the sense that a particular operator may assign a label one frame ahead of or behind that which would have been assigned by a different operator. An informal
comparison of the coding techniques of several trained operators showed that label assignments rarely differed by more than one frame.

**Saccade Size**

The size of a saccade is determined by measuring the amount that the eye has rotated from one fixation to the succeeding fixation. The measurement is in units of degrees of visual angle. An eye movement is indicated by movement of the cursor that is superimposed on the scene videotape to indicate direction of gaze.

Gaze shift is not solely the result of a saccadic eye movement, but is also due to movement of the head, since gaze change is the summation of eye-in-head and head-in-space coordinates. Saccade size, however, refers to the actual rotation of the eyeball in the eye socket, and is what is recorded during the coding process.

Saccade size was found by measuring the displacement of the cursor between successive fixations using a specially calibrated ruler. The calibrated ruler was created in the following way:

1. Two markers were placed on a wall, 48 inches apart, as shown in Figure 3-20.

![Figure 3-20. Calculation of visual angle from field of view.](image-url)
2. The scene camera from the portable eye tracker was placed 31 inches in front of the wall, and a videorecording was made of the markers on the wall.

3. The visual field of view captured by the camera was found from:

\[
\tan \theta = \frac{24}{31} = 0.774
\]

\[
\theta = 37.75^\circ
\]

\[
2\theta = 75.5^\circ = \text{field of view captured by scene camera}
\]

4. The tape was played back through the VCR, and displayed on the monitor. The width of scene on the video monitor was measured and found to be 23 cm.

5. The calibration of the video monitor to the field of view of the scene camera corresponds to:

\[
\frac{75.5^\circ}{23 \text{ cm}} = 3.28^\circ/\text{cm}, \text{or } 0.3 \text{ cm/degree visual angle}
\]

6. A ruler was created that marked off 3 degrees of visual angle every 0.9 cm.

Using the calibrated ruler, saccade size could be measured directly from the video monitor screen.

**3.4.3 Statistical Analysis**

Once the videotape has been completely coded, the information can be exported to a spreadsheet or some other type of statistical analysis software. The distribution of fixation durations and saccade sizes can be determined. Histograms of fixation durations show that they are typically exponentially distributed. According to Epelboim (1999), this type of distribution can best be described by a Poisson process, and is modeled by a gamma distribution.

A process can be assumed to be Poisson if it meets three criteria:
1. The observation period can be divided up into sufficiently small subintervals such that no more than one event could occur in that subinterval.

2. Within each subinterval, either an event occurs or it does not occur.

3. Each event can be regarded as occurring independently from any other event.

The time interval between each saccadic eye movement (fixation duration) clearly meets all three of these criteria, therefore the process can be considered Poisson. The sum of a series of Poisson probabilities is an expression of the distribution function of a gamma-distributed random variable (Wackerly, et al, 1996). A gamma distribution with parameters A and B is given by

\[
 f(y) = \frac{y^{A-1} e^{-y/B}}{B^A \Gamma(A)} , \quad A, B > 0; \quad 0 \leq y < \infty
\]

\[
 f(y) = 0 \quad \text{elsewhere}
\]

where

\[
 \Gamma(A) = \int_0^\infty y^{A-1} e^{-y} \, dy. \tag{3}
\]

\(\Gamma(A)\) is known as the gamma function, and is defined recursively as \(\Gamma(1) = 1\), and

\(\Gamma(A) = (A - 1)\Gamma(A - 1)\) for any \(A \geq 1\). Thus, if \(n\) is an integer, \(\Gamma(n) = (n-1)!\)

A is the shape parameter associated with the gamma distribution, and B is the scale parameter. Figure 3-21 shows a gamma probability distribution with \(B = 1\) and \(A = 2\).
Once the distributions of fixation durations have been found, it is usually useful to be able to compare the means for two different sets of criteria. A t-test is used to determine if the means between two different sets of data are significantly different. The t-test assumes that the distribution is normally distributed, however it has been found from empirical studies that the t-test is robust to the assumption of normality, that is, deviations from normality do not significantly effect the outcome of the t-test (Wackerly, et al, 1996).

Test statistic, \( t = \frac{\bar{y}_1 - \bar{y}_2 - D_0}{S_p \sqrt{(1/n_1 + 1/n_2)}} \)  
where \( S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \) \hspace{1cm} (3)

The terms \( \bar{y}_1 \) and \( \bar{y}_2 \) refer to the means of the first and second data sets, \( D_0 \) refers to the hypothesized difference between the two means (typically it is set to 0), \( n_1 \) and \( n_2 \) are the number of samples in the two data sets respectively, and \( S_1 \) and \( S_2 \) refer to the standard deviations of the two data sets. If the value of the t-test does not fall in the rejection region, then the null hypothesis is not rejected, and there exists insufficient evidence to indicate that the difference in the means is significant. For example, for an \( \alpha = .05 \) confidence level, a t-test indicating \( p<.05 \) means that the likelihood that the means are not significantly different is less than .05.
Chapter 4

4. **Experiment 1 — Moving Through a Hallway**

Four experiments were designed and carried out that used the portable eye-tracking equipment of the VPL. The overall goal of the experiments was to characterize the eye movements made by subjects engaged in everyday, unrestrained tasks in a natural environment, and to discover how the eye-movement data could uncover the pre-conscious strategies employed by those subjects. The tasks of the experiments are intended to be a series of progressively more complex situations that test oculomotor behavior in a variety of ways.

The first experiment consisted of monitoring subjects' eye-movements under varying conditions as they moved through a hallway. The second experiment consisted of monitoring fixation stability during self-motion. The third experiment was an investigation of eye movements made while washing one's hands. The fourth experiment was an investigation of eye movements made while making a selection from a vending machine.

The tasks required for the set of experiments were designed to be progressively more complex. One may argue that in real life the task of washing one's hands is no more
complex than that of walking through a hallway (if obstacle avoidance, etc. is to be considered), however, a task such as hand-washing affords a greater cognitive challenge than does simple movement. Hand-washing, while it may seem easier on an intuitive level because it is a common over-learned task, requires decisions to be made, strategies to be formulated, physical actions to be coordinated, and goals to be set and accomplished. Even an over-learned task such as hand-washing requires constant visual monitoring of the environment to a much greater extent than does simply moving around. In this sense a record of eye movements made while performing a task such as hand-washing can reveal much more about cognitive processes than can a simpler task such as movement. By the same token, making a selection from a vending machine adds another level of cognitive complexity in that it is largely not an over-learned task. The decision-making process is much more complex, and the action coordination must proceed from planning rather than rote behavior. The eye movement data from this type of task has the potential to reveal the complex interplay between pre-conscious thought and overt behavior.

4.1 Objective

The primary objective of Experiment 1 was to begin an investigation into eye movements in a real-world, non-laboratory setting using portable eye-tracking equipment. This was done by recording the eye-movements of subjects as they performed a series of tasks that involved either real self-motion, or self-motion that was simulated on a television monitor. Four experimental conditions were used to generate the data: Real/Active motion (actually moving around), Real/Passive motion (being pushed in a wheelchair), Simulated/Active motion (watching a videotape of someone moving around), and Simulated/Passive motion (watching a videotape of someone being pushed in a wheelchair).
The data consisted of fixation durations and saccade sizes. The experimental goal was to determine what, if any, difference exists between the four viewing conditions in terms of these metrics, and to apply the results to previously published studies concerning the relationship between eye movements and self-motion.

4.2 Experimental Design and Conditions

Four subjects performed the experiment JB, JP, MA, and RC. Three of the four were familiar with eye-tracking experiments, and all had normal vision. Each subject performed the experiment under four separate viewing conditions:

1. **Simulated/Active** – The subject sat in a chair approximately 3 feet from a 27" television monitor as shown in Figure 4-1, and watched a pre-recorded videotape of a scene depicting movement through a hallway. The display on the television screen subtended an area of 36 x 27 degrees of visual angle.

![Figure 4-1. Simulated condition set-up for Exp. 1](image)

The scene was recorded by a person walking through the hallway with a Hi-8 camcorder. The steady-cam function was turned on to eliminate the effects of arm motion during recording. The motion of the scene included the bounce and sway of natural gait at a normal pedestrian speed because the tape had been
recorded by a person actually walking through the hallway. All four conditions depicted the same scene, which consisted of four straight hallways connected in a square configuration, as shown in Figures 4-2 through 4-5.

2. **Simulated/Passive** – This condition was the same as for Simulated/Active, except the motion of the scene was smooth; it did not include the elements of bounce and sway.

3. **Real/Active** – The subject walked around the same hallway as depicted in the videotape at a normal pedestrian speed.
4. **Real/Passive** – The subject was pushed in a wheelchair around the hallway.

Eye movements were recorded for each subject while they performed the experiment under each of the four conditions. Example scenes of the recorded videotape, with the superimposed cursor depicting point of gaze and eye image in the upper right corner, are shown in Figures 4-6 through 4-9.

![Figure 4-6. Simulated/Active](image1.png)

![Figure 4-7. Simulated/Passive](image2.png)

![Figure 4-8. Real/Active](image3.png)

![Figure 4-9. Real/Passive](image4.png)
4.3 Data Analysis and Results

Fixation Durations

Figures 4-10 through 4-13 depict histograms of fixation durations for all four of the subjects under the condition of Simulated/Active motion.
Figures 4-14 through 4-17 depict histograms for fixation durations for all four of the subjects under the condition of Simulated/Passive motion.

**Figure 4-14. Fixation Durations Simulated Passive - JB**

**Figure 4-15. Fixation Durations Simulated Passive - JP**

**Figure 4-16. Fixation Durations Simulated Passive - MA**

**Figure 4-17. Fixation Durations Simulated Passive - RC**

<table>
<thead>
<tr>
<th></th>
<th>FD - SPJB</th>
<th></th>
<th>FD - SPJP</th>
<th></th>
<th>FD - SPMA</th>
<th></th>
<th>FD - SPRC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>382</td>
<td>490</td>
<td>316</td>
<td>624</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>27.2</td>
<td>51.2</td>
<td>20.1</td>
<td>63.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>234</td>
<td>301</td>
<td>233</td>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td>203</td>
<td>133</td>
<td>167</td>
<td>167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Dev</td>
<td>365</td>
<td>581</td>
<td>283</td>
<td>684</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Var</td>
<td>132966</td>
<td>337637</td>
<td>79807</td>
<td>467530</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>2101</td>
<td>3200</td>
<td>1901</td>
<td>4568</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>66</td>
<td>33</td>
<td>33</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>2167</td>
<td>3233</td>
<td>1934</td>
<td>4633</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>68700</td>
<td>63199</td>
<td>62260</td>
<td>73066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>180</td>
<td>129</td>
<td>197</td>
<td>117</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figures 4-18 through 4-21 depict histograms of fixation durations for all subjects under the condition of Real/Active motion.

**Figure 4-18. Fixation Durations**
Real Active - JB

**Figure 4-19. Fixation Durations**
Real Active - JP

**Figure 4-20. Fixation Durations**
Real Active - MA

**Figure 4-21. Fixation Durations**
Real Active - RC

---

### FD - RAJB
- **Mean**: 362
- **Standard Error**: 29.3
- **Median**: 266
- **Mode**: 100
- **Standard Deviation**: 358
- **Sample Variance**: 127924
- **Kurtosis**: 13
- **Skewness**: 3
- **Range**: 2434
- **Minimum**: 33
- **Maximum**: 2457
- **Sum**: 53954
- **Count**: 149

### FD - RAJP
- **Mean**: 292
- **Standard Error**: 22.5
- **Median**: 199
- **Mode**: 167
- **Standard Deviation**: 315
- **Sample Variance**: 99493
- **Kurtosis**: 10
- **Skewness**: 3
- **Range**: 1867
- **Minimum**: 33
- **Maximum**: 1900
- **Sum**: 57224
- **Count**: 196

### FD - RAMA
- **Mean**: 238
- **Standard Error**: 16.1
- **Median**: 167
- **Mode**: 167
- **Standard Deviation**: 236
- **Sample Variance**: 55537
- **Kurtosis**: 14
- **Skewness**: 3
- **Range**: 1867
- **Minimum**: 33
- **Maximum**: 1900
- **Sum**: 50972
- **Count**: 214

### FD - RARC
- **Mean**: 265
- **Standard Error**: 16.6
- **Median**: 167
- **Mode**: 167
- **Standard Deviation**: 266
- **Sample Variance**: 70505
- **Kurtosis**: 12
- **Skewness**: 3
- **Range**: 2000
- **Minimum**: 33
- **Maximum**: 2033
- **Sum**: 68033
- **Count**: 257
Figures 4-22 through 4-25 depict histograms of fixation durations for all subjects under the condition of Real/Passive motion.

**Figure 4-22. Fixation Durations**
Real Passive - JB

<table>
<thead>
<tr>
<th>Statistics</th>
<th>JB</th>
<th>JP</th>
<th>MA</th>
<th>RC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>370</td>
<td>221</td>
<td>253</td>
<td>264</td>
</tr>
<tr>
<td>Standard Error</td>
<td>42.6</td>
<td>17.0</td>
<td>17.0</td>
<td>16.0</td>
</tr>
<tr>
<td>Median</td>
<td>234</td>
<td>150</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Mode</td>
<td>167</td>
<td>67</td>
<td>167</td>
<td>167</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>522</td>
<td>258</td>
<td>237</td>
<td>236</td>
</tr>
<tr>
<td>Sample Var.</td>
<td>272226</td>
<td>66737</td>
<td>56234</td>
<td>55815</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>42</td>
<td>45</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.6</td>
<td>0.6</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Range</td>
<td>4767</td>
<td>2733</td>
<td>2067</td>
<td>1666</td>
</tr>
<tr>
<td>Minimum</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>Maximum</td>
<td>4800</td>
<td>2766</td>
<td>2100</td>
<td>1700</td>
</tr>
<tr>
<td>Sum</td>
<td>55429</td>
<td>51236</td>
<td>49063</td>
<td>57197</td>
</tr>
<tr>
<td>Count</td>
<td>150</td>
<td>232</td>
<td>194</td>
<td>217</td>
</tr>
</tbody>
</table>
A two-way ANOVA was performed on a segment of the data to determine if there were differences between any of the conditions over all of the subjects. 400 samples of fixation duration were taken for each condition of Real/Active, Real/Passive, Simulated/Active, and Simulated/Passive. 100 samples came from each of the four subjects, and this data was combined to create 400 samples for each condition. The results are shown below.

2 way analysis of variance for Experiment 1 – Fixation Durations
400 samples for each condition. 4 subjects - 100 samples each per condition.

<table>
<thead>
<tr>
<th>Individual 95% CI</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real/Act</td>
<td>284</td>
<td>(+-*++)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real/Pas</td>
<td>270</td>
<td>(+-*--)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim/Act</td>
<td>440</td>
<td>(+-*--)</td>
<td>(-<em>-</em>)</td>
<td></td>
</tr>
<tr>
<td>Sim/Pass</td>
<td>452</td>
<td>(+-*--)</td>
<td></td>
<td>(-<em>-</em>)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual 95% CI</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB</td>
<td>391</td>
<td>(+-*--)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JP</td>
<td>339</td>
<td>(+-*--)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>264</td>
<td>(+-*--)</td>
<td></td>
<td>(+-*--)</td>
</tr>
<tr>
<td>RC</td>
<td>453</td>
<td>(+-*--)</td>
<td></td>
<td>(+-*--)</td>
</tr>
</tbody>
</table>

As can be seen from the graphs of the variance, both of the real conditions have significantly shorter fixation durations than do either of the simulated conditions. The graphs of the subject variances show that there is also some difference between the subjects in their mean fixation duration over all four conditions.
Three of the subjects (JP, MA, and RC) showed a significant difference (p < 0.02) between the real and simulated conditions for both active and passive movement. The fixation durations were significantly shorter for the real-world conditions; this implies that the rate of saccades per second was much higher for these conditions. The largest difference was for RC (the author), who had a mean fixation duration of 624 msec for the Simulated/Passive condition, and 265 msec for the Real/Active condition. This was the extreme, and while it is possible that it was due to an unconscious bias on the part of the experimenter, all of the subjects except for JB also showed a significant difference.

Shorter fixations imply more frequent eye movements over a given period of time. This indicates that people who are actively moving through a real environment look at many more things than do people who are watching a simulation of moving through the same environment. This is somewhat surprising, given that the need to monitor one's direction of heading is more pressing in the real condition than it is in the simulated condition. Yet subjects frequently move both their head and their eyes while they are walking around and do not have trouble negotiating turns or maintaining their present direction while they are doing so. When subjects were pushed in a wheelchair, they still made many more head and eye movements than they did in either of the simulated conditions. Apparently, subjects were more interested in visually exploring a real environment during real movement, either active or passive, than they were in exploring a simulation of movement while seated. Granted, the subjects' field of view was much more limited in the simulated conditions (by the perimeter of the television screen) than in the real world conditions. However, since the goal of this experiment was to find out whether a simulated motion experiment can justifiably be substituted for a real motion experiment when studying the perception of self-motion, the limitations imposed by a constrained field of view must be considered.
Therefore, no attempt was made to simulate a real-world field of view; the subjects experienced the same sized field of view that they would encounter in a constrained laboratory experiment.

Note also that the distribution modes for all conditions are approximately the same; they were between 167 and 250 msec long. This is consistent with findings on typical fixation durations as reported in the literature review (Carpenter, 1988). However, graphs showing strings (repeated occurrences) of consecutive short fixations confirm that there are many strings of very short (< 200 msec) fixations for both the real and the simulated conditions. These are shown in Figures 4-26 through Figures 4-41.
It is surprising to find so many strings of consecutive short fixations because it has generally been accepted, as mentioned in the literature review, that the time required to plan and execute a saccade is greater than 200 msec (Carpenter, 1988). A short string of two short fixations in a row could be explained as pre-planning a sequence of saccades. However, a long string of three or more short fixations in a row is evidence against pre-planning because it would be difficult to pre-plan such a long sequence. Long sequences of short fixations could have several other explanations. One hypothesis is that they could be a sequence of random eye movements with no cognitive or perceptual relevance. Another is that they could be part of a pre-cognitive strategy that employs short fixations as a means of gathering information visually about the environment, perhaps to be used at a later time. Further testing of eye movements made during the execution of a natural task can reveal cognitive strategies and provide evidence either in support of or refuting these hypotheses.

**Saccade Sizes**

Figures 4-42 through 4-45 depict histograms of saccade size for all subjects under the Simulated/Active condition. Figures 4-46 through 4-49 depict histograms for all subjects under the Simulated/Passive condition. Figures 4-50 through 4-53 depict histograms for all subjects under the Real/Active condition, and Figures 4-54 through 4-57 depict histograms for all subjects under the Real/Passive condition. The mean saccade size for all subjects was significantly larger (p < 0.05) for the real conditions than it was for the simulated conditions. The mean of all subjects was 9 degrees visual angle for the real conditions, and 6 degrees for the simulated conditions. The difference is most likely due to the experimental conditions—the perimeter of the television screen bounded the viewing area of the simulated conditions.
### SS - SAJB

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7</td>
</tr>
<tr>
<td>Standard Err</td>
<td>0.3</td>
</tr>
<tr>
<td>Median</td>
<td>6</td>
</tr>
<tr>
<td>Mode</td>
<td>6</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>4</td>
</tr>
<tr>
<td>Sample Var</td>
<td>14</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2</td>
</tr>
<tr>
<td>Skewness</td>
<td>1</td>
</tr>
<tr>
<td>Range</td>
<td>21</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>22</td>
</tr>
<tr>
<td>Sum</td>
<td>1220</td>
</tr>
<tr>
<td>Count</td>
<td>174</td>
</tr>
</tbody>
</table>

### SS - SAJP

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6</td>
</tr>
<tr>
<td>Standard Err</td>
<td>0.4</td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
</tr>
<tr>
<td>Mode</td>
<td>3</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>5</td>
</tr>
<tr>
<td>Sample Var</td>
<td>26</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2</td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
</tr>
<tr>
<td>Range</td>
<td>22</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>23</td>
</tr>
<tr>
<td>Sum</td>
<td>946</td>
</tr>
<tr>
<td>Count</td>
<td>150</td>
</tr>
</tbody>
</table>

### SS - SAMA

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7</td>
</tr>
<tr>
<td>Standard Err</td>
<td>0.3</td>
</tr>
<tr>
<td>Median</td>
<td>5</td>
</tr>
<tr>
<td>Mode</td>
<td>3</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>4</td>
</tr>
<tr>
<td>Sample Var</td>
<td>20</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2</td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
</tr>
<tr>
<td>Range</td>
<td>23</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>24</td>
</tr>
<tr>
<td>Sum</td>
<td>1678</td>
</tr>
<tr>
<td>Count</td>
<td>249</td>
</tr>
</tbody>
</table>

### SS - SARC

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4</td>
</tr>
<tr>
<td>Standard Err</td>
<td>0.2</td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
</tr>
<tr>
<td>Mode</td>
<td>3</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>2</td>
</tr>
<tr>
<td>Sample Var</td>
<td>4</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1</td>
</tr>
<tr>
<td>Skewness</td>
<td>1</td>
</tr>
<tr>
<td>Range</td>
<td>10</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>11</td>
</tr>
<tr>
<td>Sum</td>
<td>611</td>
</tr>
<tr>
<td>Count</td>
<td>143</td>
</tr>
</tbody>
</table>
Figure 4-50. Saccade Size
Real Active - JB

Figure 4-51. Saccade Size
Real Active - JP

Figure 4-52. Saccade Size
Real Active - MA

Figure 4-53. Saccade Size
Real Active - RC

<table>
<thead>
<tr>
<th>SS - RAJB</th>
<th>Mean</th>
<th>Standard Error</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Deviation</th>
<th>Sample Variance</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>0.5</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>33</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>2</td>
<td>26</td>
<td>1013</td>
<td>118</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS - RAJP</th>
<th>Mean</th>
<th>Standard Error</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Deviation</th>
<th>Sample Variance</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>0.6</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>50</td>
<td>-1</td>
<td>1</td>
<td>26</td>
<td>1</td>
<td>27</td>
<td>1612</td>
<td>162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS - RAMA</th>
<th>Mean</th>
<th>Standard Error</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Deviation</th>
<th>Sample Variance</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>0.5</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>37</td>
<td>2</td>
<td>1</td>
<td>32</td>
<td>1</td>
<td>33</td>
<td>1478</td>
<td>166</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS - RARC</th>
<th>Mean</th>
<th>Standard Error</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Deviation</th>
<th>Sample Variance</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>0.4</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>28</td>
<td>3</td>
<td>2</td>
<td>31</td>
<td>2</td>
<td>33</td>
<td>1669</td>
<td>223</td>
</tr>
</tbody>
</table>
As was done for fixation durations, a two-way ANOVA was performed on a segment of the data to determine if there were any differences in saccade size between any of the conditions over all of the subjects. Again, 400 samples of saccade size were taken for each condition of Real/Active, Real/Passive, Simulated/Active and Simulated/Passive. 100 samples came from each of the four subjects, and this data was combined to create 400 samples for each condition. The results are shown below.

2 way analysis of variance for Experiment 1 – Saccade Size
400 samples for each condition. 4 subjects – 100 samples each per condition.

<table>
<thead>
<tr>
<th>Individual 95% CI</th>
<th>condition</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real/Act</td>
<td>8.48</td>
</tr>
<tr>
<td></td>
<td>Real/Pas</td>
<td>8.78</td>
</tr>
<tr>
<td></td>
<td>Sim/Act</td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td>Sim/Pass</td>
<td>6.38</td>
</tr>
</tbody>
</table>

Subject variances show that there is also some difference between the subjects in their mean saccade size over all four conditions.
The differing amount of viewing area available to the subject for the real vs. the simulated conditions was a deliberate aspect of the experimental design. It was included to find out whether the limitations imposed by a constricted viewing area would affect oculomotor behavior. From the histograms on saccade size, it was found that eye movements made during active exploration of the real world is quantitatively different from those made while viewing a realistic simulation of the world, as displayed on a television monitor. For practical reasons, most of the previous studies that have been conducted on the relationship between eye movements and the perception of self-motion have used the artificial, simulated environment. Therefore it is useful to know what, if any, differences exist between the real and simulated environments. This experiment has shown that saccade sizes are larger and fixations are shorter in the real world than they are in the simulated world.

Figures 4-58 through 4-73 depict scatter plots of saccade size vs. fixation durations for all of the subjects over all four of the conditions. The saccade associated with each fixation is the one that immediately follows the fixation. The purpose of these plots is to determine whether short fixations are correlated with short, corrective saccades, or if the short fixations are more broadly distributed over the saccade size. As the plots show, for virtually every subject over every condition, the short fixations were broadly distributed over saccade size. The only exception was for RC (the experimenter) for the two simulated conditions. For these cases, all of the saccades for both the long and short fixations were short (less than 15 degrees visual angle).
4.4 Conclusion

From this data, short fixations are not generally associated with short, corrective saccades, but are widely distributed. This is especially true in the real-world conditions. This is evidence in support of the hypothesis that short fixations are neither random nor perceptually irrelevant, but rather serve an important perceptual function. One possibility is that they aid in the formulation of strategies for planned actions. This possibility is explored further in subsequent experiments as part of this research effort.
Chapter 5

5. Experiment 2 – Fixation Stability

5.1 Objective

The purpose of Experiment 2 was to discover how well subjects are able to hold fixation on a series of targets under three separate conditions. For the first condition, the subject watched a videotaped simulation of walking through a hallway that was displayed on a small (27") television monitor. This monitor was the same one as was used for Experiment 1. For the second condition, the subject watched the same videotaped simulation of walking, but the videotape was displayed on a large screen. For the third condition, the subject actually walked through the hallway that was depicted in the simulation. For each condition the subject was asked to look at a series of targets in the environment. The conditions were set up so that the optic flow to the subject was very nearly the same for all three conditions.

The objective was to create three situations where optic flow could be used as a cue for maintaining fixation, and to determine if that cue was useful. If subjects cannot maintain fixation very well for a particular condition, then optic flow is not useful for maintaining fixation for that condition.

Optic flow refers to the changing pattern of light that reaches the eye during movement, after having been reflected off surfaces in the environment. Flow can result
from either self-motion, motion in the environment, or from a combination of the two. When people walk through a static environment, most of the flow field is due to their own forward motion, and also to motion from bounce and sway as they walk.

Bounce and sway refers to the vertical and horizontal oscillatory movements of the body from a person’s gait as they walk. We are not usually aware of the effect of this motion on our ability to fixate objects while we are walking, however, a videotape made by someone who is walking around shows the effect quite dramatically. The simulations of Conditions 1 and 2 capture the effect of bounce and sway on the videotape, and present it to the subject to determine their fixation stability in the presence of this type of motion, but without the usual physical feedback of self-motion due to proprioception.

Retinal flow refers to motion of the image that is projected onto the retina, and is composed of both optic flow coming from the environment and motion due to the rotation of the eye during an eye movement. If a person is looking straight ahead and does not make any eye movements, retinal flow and optic flow will be the same. However, if an eye movement is made, then the optic flow will be the same as it was for looking straight ahead, but the retinal flow will be different. In order to maintain fixation on objects in the environment while a person is walking, the person must move his or her eyes to counter the effect of bounce and sway. Thus, it would seem logical to conclude that retinal flow, rather than optical flow, is the more important cue for motion. The degree to which a person must move his or her eyes while walking in order to maintain fixation on a fixed target in the environment is an indication of how important the retinal cue is.

5.2 Experimental Design and Conditions

Two subjects participated in the experiment: JB and JK. The experiment consisted of three tasks:
1. The subject watched a videotaped simulation of walking through a hallway while seated 2 ½ feet in front of a small (27 inch) television monitor showing the videotape, as shown in Figure 5-1. The monitor encompassed a 36° visual angle horizontal and 27° vertical field of view for the subject. The edges of the monitor were clearly visible to the subject during the experiment. The videotape was captured by an individual walking through the hallway recording the scene on a camcorder. The camcorder was set with the "steady-shot" option turned on to eliminate the effect of arm and hand motion during recording.

2. The subject watched the same videotape while standing approximately 8 feet in front of a large (13 feet wide by 8 ½ feet high) screen showing the same videotape, as shown in Figure 5-2. The screen encompassed 78° visual angle horizontally and 56° vertically. The edges of the screen were clearly visible to the subject during the experiment.

3. The subject walked through the hallway that was depicted in the videotape, as shown in Figure 5-3.

Figure 5-1. Set-up for Condition 1: Small Screen
The subjects' eye movements were monitored and recorded using the VPL portable eye-tracking hardware and ASL software for all three conditions. Subject JB performed Condition 2 first, then Condition 1, and finally Condition 3. Subject JK performed Condition 1 first, then Condition 2, and finally Condition 3.

The subjects were instructed to hold fixation as well as they were able to on a series of 3½ inch diameter red circular targets that were placed along a path in the environment on white walls at approximately eye height. Ten targets were used for the entire experiment, but
only the first two that were encountered by the subject were analyzed for fixation stability. The first target was located to the right of the subject's direction of heading (which was straight ahead), as shown in Figure 5-3, at an initial distance of 27 feet from the subject. The second target was located straight ahead of the subject at a distance of 58 feet from the first target. Subjects were instructed to maintain fixation on each target until the target went out of view, and then to immediately shift their gaze to the next target. Subjects were familiar with the location of each target before the experiment began to minimize the time and effort involved in searching for each target.

5.3 Data Analysis and Results

Fixation stability was measured in the vertical and horizontal directions for all three conditions. The measurements are in units of degrees of visual angle between the center of the target and the center of the cursor indicating point of gaze. A negative value for visual angle indicates that gaze was to the left (horizontal) or below (vertical) the center of the target. Data was plotted as degrees visual angle deviation from the center of the target in the horizontal direction vs. the vertical direction. Data was taken from the initial fixation on the target until the target went out of view. Figures 5-4 and 5-5 show the results for subject JB. Figure 5-4 shows the fixation stability for the target on the right and Figure 5-5 shows the same for the target straight ahead, for all three conditions of small screen, large screen, and real walking.

JB clearly showed a higher degree of fixation stability for the real walking condition in both the horizontal and vertical directions. There was little difference between the large and small screen conditions for this subject. There was also no significant difference in fixation stability for the target on the right as compared to the target straight ahead. It is interesting to note that fixation stability for the real walking condition, while better than that for the
simulated conditions, was far from perfect. Much of the time the target was not centrally located on the fovea.

Figure 5-4. Fixation stability for target on right for subject JB

Figure 5-5. Fixation stability for target straight ahead for subject JB
Figures 5-6 and 5-7 show the same information for subject JK. For JK, the vertical component of fixation stability for the real walking condition was much less stable than for the horizontal component. This could be due to tracking error from the eye-tracking equipment or a calibration error, and may not indicate real fixation instability. Also note that the fixation stability for the real walking condition is about the same as for the small screen condition, and better than for the large screen condition. This is the case for both the target straight ahead and the target on the right, and for both horizontal and vertical directions.

Figures 5-8 through Figures 5-10 show a series of three frames for each of the conditions taken from the videotape of the experiment for subject JB. The series represents the three most deviant positions of the cursor from the center of the target for that particular condition, and are a visual indication of the amount of fixation instability.

Overall, these results indicate that fixation stability is most accurate for subjects who are actively moving in their environment, yet the stability is not as accurate as one would expect. Subjects were able to tolerate a reasonable amount of retinal image instability in the real world condition, as it did not impair their ability to move around. The large amount of retinal image motion experienced in the simulated conditions was too much for the subjects to bear, and greatly impeded the subjects' perception of self-motion. This finding is in general agreement with Collewijn, et al. (1992) who found that removing the head from the bite-bar destabilized fixation stability, yet did not impede subjects' accuracy of saccadic landing.
Figure 5-6. Fixation stability for target on right for subject JK

Figure 5-7. Fixation stability for target straight ahead for subject JK
Figure 5-8. Fixation stability for small screen condition. The dot indicates the target, and the square indicates the point of gaze. Subject JB.

Figure 5-9. Fixation stability for large screen condition. The dot indicates the target, and the square indicates the point of gaze. Subject JB.

Figure 5-10. Fixation stability for real walking condition. The dot indicates the target, and the square indicates the point of gaze. Subject JB.
The data indicates that fixation is more easily maintained during real motion than during passive viewing of simulated motion, although fixation is never perfect even for real motion. Fixation stability also becomes worse as the size of the simulated environment increases. This indication is supported by verbal reports given by the subjects after the experiment had ended. This is also obvious to anyone who has ever watched a videotaped scene that had been recorded by someone holding the camcorder as they moved around. The apparent motion of the environment due to the bounce and sway of the cameraperson can sometimes be enough to make people sick, yet it is the same optic flow available to someone moving around in the world.

Retinal flow, as suggested by Cutting, et al. (1992) is composed of optic flow, rotational flow due to eye movements, and translational flow from bounce and sway. The information from retinal flow should be sufficient to compensate for the apparent motion of the environment. This experiment has shown that this is not the case for subjects who are actively moving around in their environment. The perception of a stable world persists despite the effects of a considerable amount of motion on the retina, with or without eye movements.

Extra-retinal proprioceptive and vestibular signals are available during active movement that are not available during passive viewing of simulated motion. These signals enable a near real-time counter-rotational movement of the eye to offset the apparent motion of the environment due to bounce and sway. This is not a conscious eye movement, but rather one that is programmed as part of the VOR to allow for fixation stability.

Simulated self-motion presents a different kind of challenge to the oculomotor system. Slower pursuit eye movements are necessary to maintain a foveated image when the
environment appears to be moving but the individual is not. Prediction is difficult because of the lack of extra-retinal signals.

5.4 Conclusion

The conclusion of this experiment is that it is neither pure retinal flow nor pure optical that is the source of information for maintaining fixation on objects as one moves about. Rather, it is a combination of the flow field (either retinal or optical) and the context of the environment in which the motion is taking place that enables a perception of stability.

If the environment is artificial or simulated, fixation stability will be hampered not only by the lack of proprioceptive and vestibular signals, but also by the lack of a "world-as-anchor" perception of continuity. In a real environment, fixation stability is far from perfect, but the individual still has the perception of stability. Thus, the environment itself acts as an external source of information to compensate for an incomplete compensatory mechanism. The context of the environment, rather than being considered as an adjunct to visual processing capabilities, is central to a complete understanding of how individuals use visual cues to interact with that environment, and enables a means of formulating strategies to perform natural tasks.
Chapter 6

6. Experiment 3 – Handwashing

6.1 Objective

The purpose of Experiment 3 was to characterize the eye movements made by individuals engaged in a natural, unrestrained, everyday task. The task chosen for this experiment was that of washing one's hands in a campus washroom.

6.2 Experimental Design and Conditions

Four subjects were used for this experiment, two male – JB and JP, and two female – AS and MH. The two men used the men's washroom, and the two women used the women's washroom. Both washrooms were similarly appointed with a left and a right sink, a soap dispenser to the upper right of each sink, a mirror, a paper towel dispenser, two stalls, and a garbage can. In the men's washroom, shown in Figure 6-1, the paper towel dispenser was located to the left of both sinks, whereas in the women's washroom, shown in Figure 6-2, it was located between the two sinks. Also, the garbage can in the men's washroom was located to the left of the paper towel dispenser, whereas in the women's washroom the garbage can was located on the opposite side of the room from the sinks.
The subjects were instructed to go to the washroom, wash their hands, and return.  
No other person was in the washroom at the time the subjects performed the task.

**Figure 6-1. Men’s washroom used for Exp. 3**  
**Figure 6-2. Women’s washroom used for Exp. 3**

### 6.3 Data Analysis and Results

The subjects' eye movements were monitored and recorded during the entire task. The data was analyzed from the time the subjects first located one of the sinks and began approaching it, until they opened the door to exit from the washroom.

#### 6.3.1 Fixation Durations

Figures 6-3 through 6-6 show the distribution of fixation durations for all four subjects for the entire task. Figures 6-7 through 6-10 show the same data, excluding the fixations made while subjects actually washed their hands in the sink. The reason for excluding this data was to determine whether the fixations made while the subjects were actually washing their hands skewed the distribution in the direction of longer fixations. In three of the four subjects (JB, JP, and MH) this was found to be the case, especially for JB, where the difference in the means was found to be 81 msec (390 msec with washing the hands, and 309 msec without).
### Figure 6-3. Fixation Durations
Handwashing - Subject AS

<table>
<thead>
<tr>
<th>Fixation Durations (msec)</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>0.00</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
</tr>
<tr>
<td>167</td>
<td>0.20</td>
</tr>
<tr>
<td>233</td>
<td>0.30</td>
</tr>
<tr>
<td>300</td>
<td>0.40</td>
</tr>
<tr>
<td>367</td>
<td>0.50</td>
</tr>
<tr>
<td>433</td>
<td>0.60</td>
</tr>
<tr>
<td>500</td>
<td>0.70</td>
</tr>
<tr>
<td>567</td>
<td>0.80</td>
</tr>
<tr>
<td>633</td>
<td>0.90</td>
</tr>
<tr>
<td>700</td>
<td>More</td>
</tr>
</tbody>
</table>

Mean = 274
Standard Err = 27.6
Median = 200
Mode = 133
Standard Dev = 231
Sample Varic = 53238
Kurtosis = 4
Skewness = 2
Range = 1034
Minimum = 66
Maximum = 1100
Sum = 19222
Count = 70

### Figure 6-4. Fixation Durations
Handwashing - Subject JB

<table>
<thead>
<tr>
<th>Fixation Durations (msec)</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>0.00</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
</tr>
<tr>
<td>167</td>
<td>0.20</td>
</tr>
<tr>
<td>233</td>
<td>0.30</td>
</tr>
<tr>
<td>300</td>
<td>0.40</td>
</tr>
<tr>
<td>367</td>
<td>0.50</td>
</tr>
<tr>
<td>433</td>
<td>0.60</td>
</tr>
<tr>
<td>500</td>
<td>0.70</td>
</tr>
<tr>
<td>567</td>
<td>0.80</td>
</tr>
<tr>
<td>633</td>
<td>0.90</td>
</tr>
<tr>
<td>700</td>
<td>More</td>
</tr>
</tbody>
</table>

Mean = 390
Standard Err = 50.9
Median = 200
Mode = 167
Standard Dev = 438
Sample Varic = 191930
Kurtosis = 4
Skewness = 2
Range = 2000
Minimum = 33
Maximum = 2033
Sum = 28858
Count = 74

### Figure 6-5. Fixation Durations
Handwashing - Subject JP

<table>
<thead>
<tr>
<th>Fixation Durations (msec)</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>0.00</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
</tr>
<tr>
<td>167</td>
<td>0.20</td>
</tr>
<tr>
<td>233</td>
<td>0.30</td>
</tr>
<tr>
<td>300</td>
<td>0.40</td>
</tr>
<tr>
<td>367</td>
<td>0.50</td>
</tr>
<tr>
<td>433</td>
<td>0.60</td>
</tr>
<tr>
<td>500</td>
<td>0.70</td>
</tr>
<tr>
<td>567</td>
<td>0.80</td>
</tr>
<tr>
<td>633</td>
<td>0.90</td>
</tr>
<tr>
<td>700</td>
<td>More</td>
</tr>
</tbody>
</table>

Mean = 280
Standard Err = 42.4
Median = 167
Mode = 67
Standard Dev = 362
Sample Varic = 131100
Kurtosis = 15
Skewness = 3
Range = 2334
Minimum = 33
Maximum = 2367
Sum = 20435
Count = 74

### Figure 6-6. Fixation Durations
Handwashing - Subject MH

<table>
<thead>
<tr>
<th>Fixation Durations (msec)</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>0.00</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
</tr>
<tr>
<td>167</td>
<td>0.20</td>
</tr>
<tr>
<td>233</td>
<td>0.30</td>
</tr>
<tr>
<td>300</td>
<td>0.40</td>
</tr>
<tr>
<td>367</td>
<td>0.50</td>
</tr>
<tr>
<td>433</td>
<td>0.60</td>
</tr>
<tr>
<td>500</td>
<td>0.70</td>
</tr>
<tr>
<td>567</td>
<td>0.80</td>
</tr>
<tr>
<td>633</td>
<td>0.90</td>
</tr>
<tr>
<td>700</td>
<td>More</td>
</tr>
</tbody>
</table>

Mean = 326
Standard Err = 59.3
Median = 167
Mode = 167
Standard Dev = 520
Sample Varic = 270720
Kurtosis = 46
Skewness = 6
Range = 4265
Minimum = 33
Maximum = 4300
Sum = 25129
Count = 77
Figure 6-7. Fixation Durations - Not Including Wash Hands
Handwashing - Subject AS

Figure 6-8. Fixation Durations - Not Including Wash Hands
Handwashing - Subject JB

Figure 6-9. Fixation Durations - Not Including Wash Hands
Handwashing - Subject JP

Figure 6-10. Fixation Durations - Not Including Wash Hands
Handwashing - Subject MH

FD minus HAND/AS

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>279</td>
</tr>
<tr>
<td>Standard Err</td>
<td>32.9</td>
</tr>
<tr>
<td>Median</td>
<td>200</td>
</tr>
<tr>
<td>Mode</td>
<td>100</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>233</td>
</tr>
<tr>
<td>Sample Varic</td>
<td>54156</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4</td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
</tr>
<tr>
<td>Range</td>
<td>1034</td>
</tr>
<tr>
<td>Minimum</td>
<td>66</td>
</tr>
<tr>
<td>Maximum</td>
<td>1100</td>
</tr>
<tr>
<td>Sum</td>
<td>13935</td>
</tr>
<tr>
<td>Count</td>
<td>50</td>
</tr>
</tbody>
</table>

FD minus HAND/JB

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>309</td>
</tr>
<tr>
<td>Standard Err</td>
<td>42.3</td>
</tr>
<tr>
<td>Median</td>
<td>200</td>
</tr>
<tr>
<td>Mode</td>
<td>100</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>314</td>
</tr>
<tr>
<td>Sample Varic</td>
<td>98525</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4</td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
</tr>
<tr>
<td>Range</td>
<td>1333</td>
</tr>
<tr>
<td>Minimum</td>
<td>33</td>
</tr>
<tr>
<td>Maximum</td>
<td>1356</td>
</tr>
<tr>
<td>Sum</td>
<td>16992</td>
</tr>
<tr>
<td>Count</td>
<td>55</td>
</tr>
</tbody>
</table>

FD minus HAND/JP

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>242</td>
</tr>
<tr>
<td>Standard Err</td>
<td>31.8</td>
</tr>
<tr>
<td>Median</td>
<td>167</td>
</tr>
<tr>
<td>Mode</td>
<td>67</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>261</td>
</tr>
<tr>
<td>Sample Varic</td>
<td>67911</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5</td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
</tr>
<tr>
<td>Range</td>
<td>1134</td>
</tr>
<tr>
<td>Minimum</td>
<td>33</td>
</tr>
<tr>
<td>Maximum</td>
<td>1167</td>
</tr>
<tr>
<td>Sum</td>
<td>16235</td>
</tr>
<tr>
<td>Count</td>
<td>67</td>
</tr>
</tbody>
</table>

FD minus HAND/MH

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>267</td>
</tr>
<tr>
<td>Standard Err</td>
<td>29.2</td>
</tr>
<tr>
<td>Median</td>
<td>167</td>
</tr>
<tr>
<td>Mode</td>
<td>166</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>243</td>
</tr>
<tr>
<td>Sample Varic</td>
<td>58957</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7</td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
</tr>
<tr>
<td>Range</td>
<td>1333</td>
</tr>
<tr>
<td>Minimum</td>
<td>33</td>
</tr>
<tr>
<td>Maximum</td>
<td>1366</td>
</tr>
<tr>
<td>Sum</td>
<td>18395</td>
</tr>
<tr>
<td>Count</td>
<td>69</td>
</tr>
</tbody>
</table>
The shapes of the curves that include the fixations made while actually washing the hands are remarkably similar to one another for all four subjects. They also all have a similar number of fixations (70-77), however their means differ widely – from 274 msec to 390 msec. This can be attributed in part to several long fixations made while actually washing the hands for subjects MH and JB. When these long fixations are excluded from the analysis, the shapes of the distributions are still similar, but the means are lower and the means also show less deviation from one another, ranging from 242 msec to 309 msec.

6.3.2 Saccade Size

Figures 6-11 through 6-14 show the distribution of saccade sizes for all four subjects. Figures 6-15 through 6-18 show the saccade sizes not including the fixations made while actually washing the hands. Unlike fixation durations, there is virtually no difference in the means for each subject when washing of the hands is excluded from the analysis. The average number of saccades is fewer, 52 over all subjects when excluding washing the hands versus 65 without. For two subjects (AS and JB) the number of saccades decreased by 29%, but it is interesting to note that the mean did not change (9° for JB) or barely changed (from 16° to 15° for AS). This is an indication that mean saccade size is not particularly sensitive to the sub-task being performed, at least for the overall task of washing one’s hands in a public restroom.

The range of saccade sizes is quite large for all four subjects, from 1° to 26° for JB and from 3° to over 30° for the other subjects. The mode is 3° for three of the subjects, and 9° for the other. This suggests that even though the mean saccade size and the mode of the saccade size does not differ much between sub-tasks, each sub-task encompasses wide variability in the amount the eyes move in order to accomplish that specific task. This is
Figure 6-11. Saccade Size
Handwashing - Subject AS

Mean 15
Standard Errc 1.1
Median 14
Mode 9
Standard Dev 9
Sample Varic 80
Kurtosis 0
Skewness 1
Range 39
Minimum 3
Maximum 42
Sum 990
Count 65

Figure 6-12. Saccade Size
Handwashing - Subject JB

Mean 9
Standard Err 0.8
Median 7
Mode 3
Standard Dev 6
Sample Varic 41
Kurtosis 1
Skewness 1
Range 25
Minimum 1
Maximum 26
Sum 583
Count 65

Figure 6-13. Saccade Size
Handwashing - Subject JP

Mean 10
Standard Errc 0.9
Median 7
Mode 3
Standard Dev 7
Sample Varic 56
Kurtosis 2
Skewness 1
Range 34
Minimum 2
Maximum 36
Sum 642
Count 65

Figure 6-14. Saccade Size
Handwashing - Subject MH

Mean 12
Standard Errc 1.1
Median 9
Mode 3
Standard Dev 9
Sample Varic 73
Kurtosis 4
Skewness 2
Range 42
Minimum 3
Maximum 45
Sum 743
Count 63
Figure 6-15. Saccade Size - Not Including Wash Hands
Handwashing - Subject AS

Figure 6-16. Saccade Size - Not Including Wash Hands
Handwashing - Subject JB

Figure 6-17. Saccade Size - Not Including Wash Hands
Handwashing - Subject JP

Figure 6-18. Saccade Size - Not Including Wash Hands
Handwashing - Subject MH

SS minus HAND/AS
Mean 16
Standard Err 1.4
Median 13
Mode 9
Standard Dev 9
Sample Varic 89
Kurtosis 0
Skewness 0
Range 39
Minimum 3
Maximum 42
Sum 720
Count 46

SS minus HAND/JB
Mean 9
Standard Err 0.9
Median 7
Mode 3
Standard Dev 7
Sample Varic 42
Kurtosis 0
Skewness 1
Range 24
Minimum 1
Maximum 25
Sum 440
Count 48

SS minus HAND/JP
Mean 10
Standard Err 1.0
Median 8
Mode 3
Standard Dev 8
Sample Varic 60
Kurtosis 2
Skewness 1
Range 34
Minimum 2
Maximum 36
Sum 587
Count 58

SS minus HAND/MH
Mean 12
Standard Err 1.2
Median 10
Mode 3
Standard Dev 9
Sample Varic 79
Kurtosis 4
Skewness 2
Range 42
Minimum 3
Maximum 45
Sum 676
Count 56
evidence that washing one's hands is a complex natural task that requires strategic planning in order to maximize the amount of visual information extracted from the environment.

6.3.3 Major Sub-tasks

As mentioned earlier, the overall task of washing one's hands can be thought of as being composed of many smaller sequentially performed sub-tasks. Figure 6-19 through 6-22 show the duration and associated label of each of these sub-tasks for the four subjects. While there exist many similarities between the subjects, each subject has his or her own idiosyncratic way of accomplishing the goal of washing one's hands. For example, when to get the soap, and whether or not to get a second paper towel for drying the hands is an individual decision, and the decisions may be different on each occasion. However, these are some of the things that all of the subjects have in common for this task:

1. All of the subjects approached the sink right away after they had entered the bathroom.

2. All of the subjects got soap before they rubbed their hands together, however, two of the subjects, JP and AS, wet their hands before they got the soap.

3. All subjects performed the sub-tasks in a particular sequence: Wash Hands, Turn Off Water, Get Paper Towel, Dry Hands, Throw Away Paper Towel. This sequence is followed by either Get Second Paper Towel, Dry Hands Again, Throw Away Second Paper Towel, Exit Bathroom, or just Exit Bathroom.

This analysis is included because it provides a framework for describing eye movements in terms of their association with certain sub-tasks. Figure 6-23 through 6-26 show a hierarchical timeline of the Handwashing sub-tasks. Level 1 corresponds to the overall goal of washing one's hands, Level 2 corresponds to the major sub-tasks as described
Figure 6-23. Handwashing Subject - AS

Figure 6-24. Handwashing Subject - JB

Figure 6-25. Handwashing Subject - JP

Figure 6-26. Handwashing Subject - MH
above, and Level 3 corresponds to object manipulations. Manipulations begin the moment a
subject makes physical contact with an object in the environment. The length of the bars at
each level indicates the amount of time that has elapsed from the start of one manipulation
to the start of the next manipulation. Level 4 corresponds to fixations.

Taken together, all four levels represent a hierarchy of goals. The goals become less
abstract as the levels are descended from Level 1 to Level 4, and increase in frequency per
unit time. The ability of subjects to consciously monitor each of these goals also decreases
as the levels are descended. According to this model, fixation durations, which occur at the
lowest level of this hierarchy, can be thought of as explicit manifestations of the higher level
goals. They are micro-tasks that need to be performed in a specific sequence in order to
support the goal of individual manipulations, which in turn need to support one of the major
sub-tasks, which finally make up the overall goal of washing one's hands. As an analogy to
machine language instruction parsing, eye movements correspond to the indivisible atomic
operations that are strung together to form the basic instruction, which are in turn compiled
into meaningful statements, the sequence of which comprises a complete program.

6.3.4 "Look-heads"

One advantage of constructing a model of eye movements that is based on a
hierarchy of goals is that the temporal relationship between specific sub-tasks can be
established. Figures 6-27 through 6-30 show the amount of time that has elapsed between a
fixation on a particular object and the physical manipulation of that object. All four subjects
showed some fixations that exhibited significant delays between the object fixation and the
object manipulation. The amount of that delay, as indicated by the height of the bar, is a
Figure 6-27. Elapsed time between object fixation and object manipulation for Handwashing - Subject AS

Figure 6-28. Elapsed time between object fixation and object manipulation for Handwashing - Subject JB

Figure 6-29. Elapsed time between object fixation and object manipulation for Handwashing - Subject JP

Figure 6-30. Elapsed time between object fixation and object manipulation for Handwashing - Subject MH
quantifiable measure of "look-ahead". Some subjects, particularly MH, exhibited significant look-aheads early in the task. She looked several times at the soap dispenser and also the paper towel dispenser before she had even turned on the water to wet her hands. The other subjects exhibited fewer look-aheads that were scattered throughout the task, however each subject looked at the soap dispenser well before they had an actual need to use it.

Figures 6-31 and 6-32 show frames from the videotape of subject JB performing the hand washing experiment, which depicts one example of look-ahead. Figure 6-31 shows JB's first fixation on the garbage can (indicated by the black cursor), which occurred at a timecode of 00:55:03. Figure 6-32 shows JB's first manipulation of the garbage can, which occurred at a timecode of 01:02:07 - a time difference of 7 seconds and 4 frames (7.13 seconds). This is a modest look-ahead, in terms of the time lapse between object fixation and manipulation, yet it is still much longer than the 0.6 seconds suggested by Land (1999) as the time difference between the first object-related fixation and the first indication of manipulation.

Figure 6-31. JB's first fixation on garbage can.
Figure 6-32. JB's first manipulation of garbage can, 7.1 seconds later.
Figures 6-33 through 6-36 show another way to graphically visualize the look-ahead phenomenon. These graphs show the temporal relationship between individual fixations (Level 4) and the major sub-tasks (Level 2). Each fixation was analyzed to determine which of the major sub-tasks that the object being fixated logically belonged to. Most of the fixations correspond to the sub-task that was currently being executed, but some of the fixations clearly correspond to major sub-tasks that occurred much later in the performance of the task. These fixations are indicated on the graph by the diagonal lines that cross over the other, more vertical lines connecting Level 4 to Level 2.

How does the observance of look-ahead fit in with the concept of a hierarchical model of sub-goals for task performance? Several possibilities exist. One is that look-aheads are necessary to formulate a strategy, or plan of action, to execute the task in an efficient manner. Given that people tend to look at objects that they will eventually manipulate (Land, 1999), the look-aheads may be an advantageous way of gathering information ahead of time, and storing it for later use. This possibility is supported by the observation that the look-aheads always occur to the relevant major objects in the environment – those that would be mentioned in a verbal description of the scene, such as sink, garbage can, soap dispenser – rather than the minor less useful objects such as ceiling tiles or door hinges.

Another possibility is that the hierarchy represents a way to manage the flow of information needed to adequately plan for the future and accomplish goal-oriented behavior. For example, any given sub-task requires a minimum number of fixations to be performed adequately. These fixations can be allocated as the task unfolds, as they are needed. Some of the fixations may be extra, in the sense that they are not crucial for the adequate
Handwashing
Subject - AS

Figure 6-33.

Handwashing
Subject - JB

Figure 6-34.

Handwashing
Subject - JP

Figure 6-35.

Handwashing
Subject - MH

Figure 6-36.
execution of the task. This is reasonable to assume because eye movements are relatively inexpensive, both computationally and physically, compared to manipulations, and usually occur without conscious control. It would be an efficient strategy to use some of these extra fixations to gather information that will most likely be needed in the future. How is it possible for such a low-level mechanism that is not under conscious control to adequately predict the future? The hierarchical model incorporates the mechanism for prediction. In the model, the top-most level has manifest conscious control over the entire execution of the task and also has global access over the control of all the Level 2 sub-tasks, which in turn has access to the next lower level. Finally, that level reaches to the lowest level of eye movements and individual fixations. The top level is the control center for coordinating the sequence of operations at all levels that is necessary to efficiently accomplish the task.

6.4 Conclusion

This experiment was an investigation into the role of eye movements during the performance of an everyday, over-learned task. The task was that of washing one's hands in a public restroom. It was an interesting experiment, and one in which much was learned, some of it expected, and some of it surprising. The expected results were the distributions of fixation durations and saccade sizes, and also the observation that most fixations correspond to the object currently being manipulated, and the sub-task currently being performed. The surprising result was the observation of the look-ahead phenomenon. A hierarchical model of sub-goals is proposed to explain both the purpose of and the control mechanism for the look-aheads. More experimentation is necessary to confirm that look-aheads also occur during the execution of other natural, everyday tasks that are not necessarily over-learned.
Chapter 7

7. **Experiment 4 – Making a Selection From a Vending Machine**

7.1 **Objective**

The objective of this experiment was to characterize the eye movements made by individuals while they are performing a natural, unrestrained, everyday task. The task chosen for this experiment was that of making a selection, purchasing, and then retrieving an item from a vending machine. This task was chosen because it represents a complex task that is not routine and that required the subject to make several decisions. This is in contrast to the task of Experiment 3, washing one's hands in a restroom, which represented a ritualized, over-learned task that required fewer overt decisions, and is less complex than making a purchase from a vending machine. The phrase "complex task" is relative; thus, one goal of this experiment is to determine if the complexity of a task is evident in the eye movements made by the subjects.
7.2 Experimental Design and Conditions

Four subjects performed the experiment, two male – JT and JB, and two female – AS and MH. They were all experienced eye-tracking subjects, but none were aware of the premise of the experiment.

The instructions given to the subjects were purposefully simple and unrestricted in order to achieve the goal of observing natural, unrestrained, everyday behavior. The subjects were told to gather enough coins together to make a vending machine purchase (an unspecified amount), and then go to the machine area and make a selection of their choice, and return.

The vending machines are located in a small alcove off a main hallway in an academic building. Inside the alcove there are four vending machine, as shown in figure 7-1. Figures 7-2 through 7-5 show the individual machines: a coffee machine, a soda machine, a candy/chip machine, and a sandwich machine. They are all approximately the same size, and each has a slot for inserting coins, a slot for inserting dollar bills, a coin release button or lever, a card insert slot, item selector buttons, a coin return tray, and an item retrieval area.

Figure 7-1. The alcove with the four vending machines used in Experiment 4.
The item selector buttons for the coffee machine and candy/chip machine are in the form of an alpha-numeric keypad, whereas the selector buttons for the soda machine are in the form of a column of large buttons, each one labeled with the brand-name of the item.
offered. There are two selector buttons for the sandwich machine. When one is pressed, the item tray inside the machine rotates in the direction indicated by the button. This locates the desired item in front of a clear sliding door which is then opened to obtain the item.

The items in the candy/chip and sandwich machine are visible to the consumer during the selection process, however, the items in the soda and coffee machine are only indicated by their label on the front of the machine. The prices for the items in all of the machines are located either near the actual item, or near their label.

In order to make a purchase, each subject was required to make several decisions, such as which machine to use, which item in that machine to buy, and whether they had enough money to purchase the item they wanted. Coincidentally, each subject chose a different machine. AS chose the soda machine, JB chose the sandwich machine, JT chose the candy/chip machine, and MH chose the coffee machine. This made for a less tightly focused experiment, yet enabled an analysis based on a generic vending machine purchase, rather than one relating to a specific type of machine.

7.3 Data Analysis and Results

The subjects' eye movements were monitored and recorded during the entire task which began when a subject entered the alcove and first fixated one of the machines, and ended when he or she left the alcove with purchase in hand.

7.3.1 Fixation Durations

Figures 7-6 through 7-9 show histograms of fixation durations for each of the four subjects. The total number of fixations for each subject varied widely, from 100 for JT to 216 for MH. This is because the sequence of events necessary to complete the task is highly individual and depends upon which machine was selected. Each subject had his or her own
Figure 7-6. Fixation Duration
Vending Machine - Subject AS

Figure 7-7. Fixation Durations
Vending Machine - Subject JB

Figure 7-8. Fixation Durations
Vending Machine - Subject JT

Figure 7-9. Fixation Durations
Vending Machine - Subject MH

FD - VEND/AS
Mean 231
Standard Errc 15.4
Median 167
Mode 100
Standard Dev. 174
Sample Varic 30311
Kurtosis 2
Skewness 2
Range 767
Minimum 33
Maximum 800
Sum 29397
Count 127

FD - VEND/JB
Mean 216
Standard Errc 15.3
Median 166
Mode 133
Standard Dev. 198
Sample Varic 39267
Kurtosis 13
Skewness 3
Range 1234
Minimum 33
Maximum 1267
Sum 36269
Count 168

FD - VEND/CT
Mean 312
Standard Errc 28.6
Median 201
Mode 200
Standard Dev. 286
Sample Varic 81851
Kurtosis 4
Skewness 2
Range 1367
Minimum 33
Maximum 1400
Sum 31171
Count 100

FD - VEND/MH
Mean 293
Standard Errc 19.4
Median 200
Mode 200
Standard Dev. 284
Sample Varic 80899
Kurtosis 15
Skewness 3
Range 1934
Minimum 33
Maximum 1967
Sum 63208
Count 216
way of deciding which item to purchase, and each machine required distinctly different kinds of manipulations to complete the transaction.

The mean fixation duration varied from 216 msec for JB to 312 msec for JT. The mode was 167 msec for three of the subjects, and 200 msec for the fourth. A two-sample t-test assuming unequal variances showed no significant difference in the mean between subjects JT and MH and between subjects AS and JB. All other subject comparisons showed significant differences ($p < 0.05$). This is an indication that making a selection from a vending machine is a highly individualized complex task.

### 7.3.2 Saccade Size

Figures 7-10 through 7-13 show histograms of saccade size for each of the four subjects. The mean saccade size ranged from $9^\circ$ to $11^\circ$ visual angle, whereas the mode was $3^\circ$ for three of the subjects and $5^\circ$ for one subject. Each subject had several large saccades, greater than $30^\circ$, but most of the data was concentrated in the $2^\circ$ to $7^\circ$ range. This observation accounts for the relatively large difference between the mean and the mode. The larger saccades were mostly associated with large head movements during gaze shifts between the four machines, while the subject was deciding which machine to use, whereas the smaller saccades were associated with small head movements while the subject was deciding which item to purchase from a particular machine.

An ANOVA performed for mean saccade size differences between the subjects yielded less evidence for individualized behavior for this task. There is no statistically significant difference between mean saccade size for AS and JB, AS and JT, and JT and MH. However, there is a small difference between JB and JT ($p < 0.04$) and a significant difference between AS and MH, and JB and MH ($p < 0.006$). Thus, the average length of a saccade is more specific to a particular individual than is the average fixation duration.
Figure 7-10. Saccade Size
Vending Machine - Subject AS

Figure 7-11. Saccade Size
Vending Machine - Subject JB

Figure 7-12. Saccade Size
Vending Machine - Subject JT

Figure 7-13. Saccade Size
Vending Machine - Subject MH

<table>
<thead>
<tr>
<th>SS - VEND/AS</th>
<th>Mean</th>
<th>Standard Err</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Dev</th>
<th>Sample Varic</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS - VEND/AS</td>
<td>11</td>
<td>0.8</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>64</td>
<td>0</td>
<td>1</td>
<td>32</td>
<td>3</td>
<td>35</td>
<td>1240</td>
<td>112</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS - VEND/JB</th>
<th>Mean</th>
<th>Standard Err</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Dev</th>
<th>Sample Varic</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS - VEND/JB</td>
<td>11</td>
<td>0.6</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>56</td>
<td>1</td>
<td>1</td>
<td>34</td>
<td>2</td>
<td>36</td>
<td>1572</td>
<td>141</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS - VEND/IT</th>
<th>Mean</th>
<th>Standard Err</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Dev</th>
<th>Sample Varic</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS - VEND/IT</td>
<td>9</td>
<td>0.9</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td>73</td>
<td>2</td>
<td>2</td>
<td>35</td>
<td>2</td>
<td>37</td>
<td>891</td>
<td>96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS - VEND/MH</th>
<th>Mean</th>
<th>Standard Err</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Dev</th>
<th>Sample Varic</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS - VEND/MH</td>
<td>9</td>
<td>0.5</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>44</td>
<td>4</td>
<td>2</td>
<td>41</td>
<td>1</td>
<td>42</td>
<td>1796</td>
<td>209</td>
</tr>
</tbody>
</table>
7.3.3 Comparison of Hand-washing and Vending Machine Experiments

Three of the subjects performed both the hand-washing experiment and the vending machine experiment – AS, JB, and MH. Subject JP performed only the hand-washing experiment, and subject JT performed only the vending machine experiment. Tables 7-1 and 7-2 show the results of a two-sample t-test assuming unequal variance for subjects AS, JB, and MH comparing the hand-washing and vending machine experiments for fixation durations and saccade size.

When comparing the fixation durations across the two experiments for the same subject, the only significant difference was for subject JB. Similar tests for saccade size showed that there were significant differences between the two tasks for all three subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixation Durations</th>
<th>Saccade Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>Hand: 274 ms</td>
<td>Vend: 231 ms</td>
</tr>
<tr>
<td>JB</td>
<td>Hand: 390 ms</td>
<td>Vend: 216 ms</td>
</tr>
<tr>
<td>MH</td>
<td>Hand: 326 ms</td>
<td>Vend: 293 ms</td>
</tr>
<tr>
<td>P(T&lt;=t);</td>
<td>0.178</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table 7-1. Comparison of fixation durations for hand-washing and vending machine.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixation Durations</th>
<th>Saccade Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH</td>
<td>Hand: 12 deg</td>
<td>Vend: 9 deg</td>
</tr>
<tr>
<td>P(T&lt;=t);</td>
<td>0.002</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Table 7-2. Comparison of saccade size for hand-washing and vending machine.

Subjects AS and MH had a larger average saccade size for hand-washing than they did for the vending machine task, whereas JB had a larger average saccade size for the vending machine task. It is not possible to distinguish between the two tasks based on either fixation duration or saccade size for these three subjects.
7.3.4 Major Sub-tasks

Figures 7-14 through 7-17 show a parsing of the overall task into the major sub-tasks, one figure for each subject.
These figures show the amount of time each subject spent completing a smaller task that contributed to achieving the overall goal of making a purchase from a vending machine. These smaller tasks are called Level 2 sub-tasks, and are shown in the figures in the order in which they were performed. It is apparent from these figures that each subject followed his or her own individual script for completing the task, and the amount of time spent at each sub-task varied from one subject to another.

Some of the Level 2 sub-tasks, particularly the early ones, require decision-making. The two sub-tasks of choosing which machine to use and which item to buy from that machine required a significant amount of time for decision-making: 52% for AS, 62% for JB, 42% for JT, and 28% for MH (who chose the coffee machine and spent the most amount of time waiting for the selection to be processed). The only exception is for subject MH, who spent more time monitoring the progress of her selection than choosing an item. This subject chose to purchase coffee and had to wait the longest of all the subjects for her purchase to be processed by the machine. She was mostly idle during this wait time, and chose to spend it comparing the prices of the other items offered by the coffee machine.

JB took the longest amount of time to choose an item, 33 seconds. He also chose to purchase two items from the sandwich machine. After attempting to retrieve the second item from the machine, he discovered that he did not put enough money into the machine to make this purchase and had to put more money in. Unintentionally, this subject carried out the most elaborate script of all the subjects, yet he spent less time performing each of the individual sub-tasks than did any of the other subjects, with the exception of choosing an initial item to buy. It would be interesting to find out if this is a peculiar trait of this particular subject, or if other subjects who display a long initial decision making stage also
show shorter times for manipulations. Also, variables such as hunger level and familiarity with the machines may affect how long the subject takes to perform a sub-task.

Figures 7-18 through 7-21 show a hierarchy of the different levels of sub-tasks associated with making a purchase from a vending machine, one figure for each subject.
Level 1, the top bar, represents the overall goal, and Level 2 represents the major sub-tasks as described above. Level 3 shows the amount of time the subjects fixated a major object in the environment, and Level 4 shows the individual fixations. The difference between Level 3 and Level 4 is that several fixations on a single major object, or portion of an object (for example, the front panel of a machine or the coin return) are grouped together as a single Level 3 sub-task, whereas each individual fixation is represented in Level 4. Level 3 can be thought of as a sequence of "attentional" fixations. Thus, several actual fixations correspond to a single "attentional" fixation, several "attentional" fixations correspond to a single Level 2 sub-task, and a string of Level 2 sub-tasks are sequentially concatenated together to accomplish the overall goal.

This is a slightly different representation than what was shown for Experiment 3. The Level 3 tasks in the hierarchy for Experiment 3 represented manipulations rather than fixations. Since there were far fewer manipulations for the vending machine task than for the hand-washing task, and the manipulations occurred mostly at the end of the task, it was more appropriate to show Level 3 for Experiment 3 as manipulations, and Level 3 for Experiment 4 as "attentional" fixations.

7.3.4 "Look-aheads"

Figures 7-22 through 7-25 show the elapsed time between object fixation and object manipulation for the four subjects. There were several instances where a subject fixated an object early in the task, during one of the decision-making sub-tasks, yet the object that was fixated was not related to the sub-task currently being performed. The sub-task relevant to that particular fixation would not occur until a later time. This is what is meant by a "look-ahead" fixation, and was described earlier for the hand-washing task. They were apparent in the hand-washing task as well as in the vending machine task.
Figure 7-22. Elapsed time between object fixation and object manipulation for Vending Machine - Subject AS

Figure 7-23. Elapsed time between object fixation and object manipulation for Vending Machine - Subject JB

Figure 7-24. Elapsed time between object fixation and object manipulation for Vending Machine - Subject JT

Figure 7-25. Elapsed time between object fixation and object manipulation for Vending Machine - Subject MH
For the vending machine task, not all of the early "look-aheads" corresponded to a later manipulation; sometimes they corresponded to an intentional "look-ahead", that is, a look toward an object that might be used sometime in the future, but is not necessarily used. For example, while JB was deciding which machine he would make a purchase from, he made a saccade to the exit tray of the candy/chip machine. This information would have been useful if he had ever purchased a candy bar, but he did not; he eventually purchased milk from the sandwich machine. Since there was no actual manipulation of the candy/chip machine exit tray to extract an item, there was no elapsed time between object fixation and object manipulation, and this "look-ahead" does not show up as a data point for subject JB in Figure 7-23.

The intentional "look-aheads" might be an example of a more general notion of the concept of "look-ahead". For a task that requires a significant amount of decision-making such as making a selection from a vending machine, fixations to objects that might be needed in the future could be an efficient strategy to reduce the overall computational load. Using visual pointers in the form of cached "look-aheads" (ones that are saved for a highly likely later retrieval) would be a good strategy to employ for a non-deterministic task, that is, one whose outcome is not pre-determined. It is likely, though not certain, that the information will be needed at a later time. Therefore, a fixation to a highly likely future target ensures that the information is stored in an easily accessible memory location.

Figures 7-26 through 7-29 show an alternate way of characterizing "look-aheads" that takes into account the more general variety, including both cached and regular. They show the temporal relationship between individual fixations and Level 2 sub-tasks. This is the same visualization technique that was used to characterize the "look-aheads" of Experiment 3 and represents the same type of information.
The "look-aheads" are shown as cross-overs from the early fixations to the Level 2 sub-tasks. The robustness and generality of this type of visualization technique is apparent when considering that it is capable of adequately characterizing both the deterministic type of "look-ahead" as well as the non-deterministic, cached type.

Comparing these figures to those of hand-washing from Experiment 3, it is apparent that there are fewer instances of "look-ahead" for this task. Also, each one extends further in time, to the later Level 2 sub-tasks. Specifically, the "look-aheads" for the vending machine task were exclusively to the item exit areas of the machine, with the exception of MH, who made several early "look-aheads" to the coffee cup lids as well as to the coffee exit area. All of the subjects looked ahead very early in the task, during the decision-making process. Subject JB, who made two purchases from the same machine, looked to the exit area twice, once each time he began the item selection process.

Figures 7-30 through 7-34 show two examples of "look-aheads" for the vending machine task taken from the videotape of subjects JB and MH performing the task. Figure 7-30 shows JB looking at the exit area of the candy/chip machine at a timecode of 8:05:02. Figure 7-31 shows him acquiring his purchase from the exit area at a timecode of 8:11:26; a time difference of 6.8 seconds between visually locating the exit area and actually using it.

Figure 7-32 shows MH looking at the coffee cup lids while she inserts the coins into the coin slot, at a timecode of 10:03:27. Figure 7-33 shows MH looking at the coffee cup lids once again while she waits for her selection to be processed, at a timecode of 10:40:10. Figure 7-34 shows MH reaching for a lid to put on her coffee cup after the transaction had been completed, at a timecode of 10:56:27. The first "look-ahead" preceded manipulation by 53 seconds, and the second "look-ahead" preceded manipulation by 16.5 seconds.
Figure 7-30. JB looks at exit of candy/chip machine at timecode of 8:05:02.

Figure 7-31. JB retrieves purchase from machine at timecode of 8:11:26, 6.8 seconds later.

Figure 7-32. MH looks at coffee cup lids at a timecode of 10:03:27.

Figure 7-33. MH looks at coffee cup lids at a timecode of 10:40:10.

Figure 7-34. MH retrieves coffee cup lid at a timecode of 10:56:27, 53 seconds after first locating the lids, and 16.5 seconds after a second look to the lids.
Another way to characterize the "look-aheads" is shown in Figures 7-35 through 7-38. These figures show the same information as Figures 7-26 through 7-29, but with the added dimension of fixation durations. The time duration of each fixation is indicated by the height of the dots that are connected to one another by the thin black line; the scale is shown along the right axis. The thick black line shows the corresponding Level 2 sub-task, with the sub-task label indicated on the left axis. A spike in this line is evidence of a "look-ahead".

From the figures, it appears the shorter fixations are associated with the "look-aheads". Figures 7-39 through 7-42 show the same information for the hand-washing task, which also associates the short fixations with the "look-aheads". For both tasks, the longer fixations tend to occur at the transitions between the Level 2 sub-tasks, occurring either just before or just after the transition. It could be that longer fixations are necessary at the transitions to stabilize the visual information on the retina in preparation for the execution of a new routine. It is also possible that they are occurring at a sub-conscious level. The "look-aheads" do not tend to occur at the boundaries of two sub-tasks, but rather in the middle of a sub-task. It could be that it is more convenient to gather new information that is not crucial for performing the current sub-task in the middle of the sub-task, when the cognitive load is lighter.

Figures 7-43 through 7-46 and 7-47 through 7-50 show the relationship between saccade size and sub-task, for both the hand-washing and the vending machine tasks. In general, the "look-aheads" are associated with the larger saccades, probably because a "look-ahead" usually requires a significant gaze shift away from the area that is currently being visually engaged.
In summary, the "look-aheads" tend to be associated with shorter fixations, longer saccade sizes, and they tend to occur in the middle of a sub-task.

Figure 7-35. Fixation Durations and Subtasks
Vending Machine - Subject AS

Figure 7-36. Fixation Durations and Subtasks
Vending Machine - Subject JB

Figure 7-37. Fixation Durations and Subtasks
Vending Machine - Subject JT

Figure 7-38. Fixation Durations and Subtasks
Vending Machine - Subject MH
7.4 Conclusion

This experiment showed that information about eye movements can be used to describe how a complex natural task unfolds over time. It also showed that it may be possible to use eye movement information to distinguish between a relatively complex, decision-making task such as purchasing an item from a vending machine and a routinized, over-learned task such as washing one's hands.

It was not possible, based on the information from this experiment, to distinguish between the two tasks from an analysis of fixation durations or saccade size. It may be possible to use the extra dimension of information provided by the "look-aheads" to further characterize complex tasks, and to formulate a conceptual framework about how strategies unfold over time. Further research is needed with many more subjects before definitive conclusions about the role of the "look-aheads" can be made.
Chapter 8

8. Conclusion and Recommendations

Overall, the objectives of this research project have been met. The primary objective was to conduct an investigation into the role of eye movements in the performance of everyday tasks in a natural environment. In addition to the primary objective were three secondary objectives. The first was to describe the effect that a real, as opposed to a simulated realistic, environment has on eye movements and the perception of self-motion while walking through a hallway. The second was to extend a two-dimensional space-variant analysis of eye movements traces into the time domain, and develop a model for describing the temporal sequencing of eye movements. A final objective was to be able to use the information found from this research to make recommendations for artificial vision systems.

8.1 The Effect of a Real Environment

Experiment 1 showed that people tend to be more actively engaged in visually exploring a real environment that they are walking through than they are an artificial, simulated environment that they view on a video monitor. This conclusion is given support from the evidence gathered for fixation durations and saccade sizes for the experimental conditions. Fixation durations are significantly shorter (thus occur more frequently) and saccade sizes are significantly longer for the real-world conditions, even when the subject is
being moved passively through the real environment. Experiment 1 also showed that very short fixation durations (less than 200 msec) occur in both real and simulated conditions, but slightly more frequently in the real-world conditions. The short fixations are not associated with short, corrective saccades, but are widely distributed over saccade size.

Experiment 2 investigated fixation stability in a real vs. a simulated environment. It gave evidence that a real environment provides a context within which the perception of a stable environment can be achieved despite a sometimes significant amount of retinal image motion during active movement. A simulated environment that does not cover the entire field of view for the observer does not provide such a context, and retinal image motion becomes quite apparent and disconcerting.

Experiments 3 and 4 were attempts to bring eye-tracking out of the laboratory and into the world of natural tasks, and represented a first step away from the classical psychophysical approach of studying eye movements within the confines and control of the laboratory setting. There exists little precedence for this kind of approach, mainly because the technology that is required to perform these experiments has not existed until recently. The hardware that was developed by the Visual Perception Laboratory at RIT specifically addresses the portability concerns that are crucial for successfully studying eye movements during natural tasks. The results of these experiments, primarily in the form of histograms of saccade sizes and fixation durations, indicate that there is sufficient reason to believe that eye movements made during the performance of natural tasks are complex and are unique to the type of task being performed.
8.2 Eye Movements Extended Over Time

A spatial analysis of an eye movement trace usually emphasizes the role that eye movements have in bringing objects of interest from the periphery to the fovea for closer inspection. It is believed that a sequence of fixations across a region of space builds up the (false) perception of a high-resolution field of view everywhere, despite evidence that visual memory across fixations tends to be very low (Irwin, 1992).

Recent studies on change-blindness (O'Regan, Rensink, and Clark, 1999), exocentric frames of reference (Pelz and Hayhoe, 1995), and memory capacity for copying colored patterns of blocks (Ballard, Hayhoe, and Pelz, 1995) have shown that the perception of a high-resolution field of view is largely illusory. The visual-perceptual system prefers, at least in these cases, to maintain a limited internal representation of physical objects in the world, and uses the environment as an external source of information, using deictic strategies to access the information when it is needed. This introduces the possibility of a temporal dimension to the sequencing of eye movements that has not been adequately studied in the past. The question is not only how visual information is acquired, but also when the information is needed.

Extending the study of eye movements over the time course of a natural task was the objective of Experiments 3 and 4. These experiments showed that when eye movement traces are considered over time as well as space, cognitive strategies such as "look-ahead" become apparent. These strategies are manifest at a very low level and are not amenable to conscious control or verbalization. A model was developed that incorporated the hierarchical structure of goal-oriented tasks, with individual fixations being the lowest level of task that was considered.
The "look-ahead" phenomenon is evidence for the sequencing of fixations based on a desire to maximize the efficiency of task performance over time, rather than for the creation of the perception of a high-resolution field of view. The "look-ahead" may be just one example of the many visual strategies yet uncovered that people use to optimize their interactions with the world under conditions of limited memory representations.

8.3 Applications for Artificial Vision Systems

Developers of artificial vision systems such as robotic systems, active vision systems, and passive surveillance systems, must contend with the limitations imposed upon their designs by currently available technology. Hardware constraints include a finite memory capacity with strict cache size and speed restrictions, bandwidth limitations for networked systems, and finite processor cycle speeds. Software constraints include program development costs and time, code maintenance, error detection, security, and platform portability. These limitations underscore the need to simplify computations as much as possible, and to develop highly efficient algorithms for implementing artificial visual systems. The amount of computation that is required for large-scale video image processing coupled with the requirement of a real-time response from the processor, can make anything but the simplest visual routines prohibitive.

One way to reduce the amount of computation is to eschew an explicit and detailed representation of the environment. Evidence gathered from this research effort, and also from those mentioned in the literature review, support the hypothesis that the human visual system makes extensive use of simplifying strategies to reduce the cognitive load required for visual processing. Anthropomorphic robotic devices that mimic human visual capabilities have been created in the past that incorporated some of these simplifications (Ballard and
Brown, 1992). This area of research is in its infancy; much more research is necessary for understanding both human vision and computer image processing in order to make robust robotic vision systems tractable.

Some insights have been gleaned from this research effort that have a direct applicability to the problem of robust computer vision. These are listed as follows:

1. The cooperative relationship between vision and action should be foremost in the design. The relationship is represented by task-specific, goal-oriented behavior, the highest level of which is defined by the user of the system. The user selects the overall goal, and the selection activates a specific branch of program execution, which then becomes the driving force behind the acquisition of all visual information.

2. The repository of information is contained in the knowledge base of the system, which is part of non-volatile memory. The knowledge base can be segmented into three areas: the internal state of the system (cold, tired, hungry, etc.), the affordances of the environment (a chair is for sitting), and past learning experiences (cushioned chairs are softer than non-cushioned chairs). The knowledge base requires no visual input initially, and can be pre-programmed, but can be later modified by information acquired visually.

3. The acquisition of new information is accomplished by descending a hierarchy of goals, the highest level of which is defined by the user, and the lowest level of which controls the camera movement. The intermediate level(s) represent a specific strategy to employ in order to realize the next higher level goal. The specific strategy to be used depends upon the current state of the overall system, which is contained in the knowledge base. The advantage of this
hierarchical architecture is that once the lowest level of explicit camera movement is reached, the only information that is acquired is that which is most pertinent to achieving the overall goal.

4. Foveation should be incorporated in the camera sensor. In other words, the environment should be sampled non-uniformly, with the highest resolution in a small central area of the sensor, and resolution falling off logarithmically in the radial direction along the periphery. Such a sensor has been realized in the laboratory, and was used in combination with a Gaussian filter and interest operator (saliency mapping) for low-level image acquisition (Yamamoto, Yeshurun and Levine, 1996).

5. Foveated images are interpreted by the knowledge base and then passed up the hierarchy of goals to change the strategy if necessary. For example, a new strategy may be implemented if the knowledge base determines that a look-ahead could be used for efficiency. The look-ahead changes the flow-of-control in much the same way that a JUMP or GOTO statement would, with control being passed to a new routine.

6. Information about the environment is maintained by the knowledge base as a local structure, and is in terms of an object-centered frame of reference, that is, the coordinates of each object in the environment are dynamic on a global map, and change as the camera moves.

7. Interaction between the knowledge base and camera sensor output is frequent and handles only small amounts of information at a time. The knowledge base determines the least amount of information that is necessary to adequately perform the task, and the camera is allowed several attempts to capture that
information. The sensor output only encodes information that currently has meaning, as determined by the knowledge base.

8. The size of working memory (contained in the camera sensor, not as part of the knowledge base) is flexible, and is determined by the task being performed. Working memory is also highly volatile; the contents are replaced frequently with new information. This requires that memory is checked frequently, and that the cameras have fast servoing mechanisms, but the advantage is that complex tasks can be performed with limited memory support.

9. The learning module of the knowledge base maintains a perception of the relative complexity of a given task. For example, if the system perceives itself to be an expert at a specific task, or if the task is perceived to be simple, fixation durations are long and gaze changes are infrequent. The opposite would hold for a complex task or for a novice system. This eases the requirements on the camera servoing mechanism for all but the most complex, or novel tasks.

10. Likewise, a task that requires the use of other mechanical parts of the robot (for example a manipulation) would allow for longer fixations, and a task that requires searching for an object would prepare the servo motor for large movements. A task that requires optical character recognition (as in reading text) would have a pre-determined fixation duration and saccade size. In other words, the movements of the camera are highly dependent upon, and can be optimized for, the task to be performed.

11. Wide tolerances for inaccuracies can be incorporated into the system to enhance performance in the presence of camera motion, camera platform
motion, and saccade target inaccuracies. Performance will be enhanced with wide tolerances because the camera servoing mechanism can be made to move faster if inaccuracies are less of a concern, allowing information to be acquired as it is needed and memory to be updated frequently. Also, it is usually easier and quicker to assimilate new information into a database if that information is incomplete and has little detail. The disadvantage is that the probability that errors will occur in the interpretation of that information becomes greater as the information becomes sparser, however an intelligent knowledge base can bring other related information to bear on the situation and easily overcome this disadvantage.

Overall, the advantages of an anthropomorphic artificial visual system seem to be compelling. After all, the human visual system has evolved over millennia to optimize the acquisition of information in the presence of limited neural hardware and an imperfect sensorimotor system. The drawbacks of an anthropomorphic system are that, much like the human system, it is slow, imprecise, and inaccurate. The great advantage is that it is general-purpose. In other words, the "job gets done" despite the imperfections and engineering design flaws. For an actively exploring robot, getting the job done despite a lack of engineering finesse may turn out to be the key to robustness.
8.4 Recommendations for Future Work

The most obvious recommendation is for more subjects to be recruited and more data to be gathered, particularly for the hand-washing and vending machine tasks. This will verify the trends in the data, and confirm the presence of look-aheads in those tasks. Also, more natural tasks should be tracked to determine if look-aheads are inherent in most, or even all, natural tasks. Also, there is the possibility that look-aheads are not "real" in the sense that they are not perceptually important. They could be simply gaze-shifts to random areas of the environment that happen to coincide with objects that will be manipulated in the future. Such a possibility is unlikely; however, the possibility should be considered.

Another question pertaining to the look-ahead phenomena is how to distinguish between visual look-aheads that are guiding actions to be performed in the near future, and true look-aheads that occur a significant amount of time before a manipulation takes place. What is the cut-off in terms of time duration between the two types of looking-ahead behaviors? Land (1999) would suggest a guiding look-ahead of 0.6 seconds or less, but this has not been confirmed for all types of tasks.

Aside from look-aheads, another anomalous finding was that of very short fixation durations, on the order of 33 msec. What is the purpose of such short fixations? Are they random eye movements, or do they have perceptual significance? Do they occur more frequently with certain kinds of tasks, and do they have a particular pattern of occurrence?

Finally, it is possible, indeed likely, that many more strategies exist that enhance the efficient performance of natural tasks besides look-aheads. As more eye-tracking experiments are performed in everyday environments, these strategies will be uncovered and a clearer window into visual perception and cognition will be revealed.
Appendix

Included in this appendix is a complete coding session for one subject (MH) for the handwashing task (Experiment 3). The session began when the subject entered the washroom and first fixated one of the sinks, and ended when the subject began to open the washroom door to exit. Each fixation begins with B_FIX in the comment column and ends with E_FIX. The number following E_FIX indicates the length of the succeeding saccade in degrees of visual angle, and the tag following the saccade length indicates the target of that saccade. The fixation duration is indicated in the third column from the left (since_last) in the same row as E_FIX.

<table>
<thead>
<tr>
<th>Timecode</th>
<th>Since_start</th>
<th>Since_last</th>
<th>Timecode</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>230433</td>
<td>-200</td>
<td>-200</td>
<td>00:03:50:13</td>
<td>start</td>
</tr>
<tr>
<td>230433</td>
<td>-200</td>
<td>00:03:50:13</td>
<td>B_FIX</td>
<td>right faucet</td>
</tr>
<tr>
<td>230433</td>
<td>-200</td>
<td>00:03:50:13</td>
<td>B_FIX</td>
<td>left sink</td>
</tr>
<tr>
<td>230699</td>
<td>66</td>
<td>266</td>
<td>00:03:50:21</td>
<td>E_FIX 21</td>
</tr>
<tr>
<td>230800</td>
<td>167</td>
<td>101</td>
<td>00:03:50:24</td>
<td>left sink</td>
</tr>
<tr>
<td>231100</td>
<td>467</td>
<td>300</td>
<td>00:03:51:03</td>
<td>E_FIX 9</td>
</tr>
<tr>
<td>231166</td>
<td>533</td>
<td>66</td>
<td>00:03:51:05</td>
<td>left soap dispenser</td>
</tr>
<tr>
<td>231166</td>
<td>533</td>
<td>00:03:51:05</td>
<td>B_FIX</td>
<td>right handle right sink</td>
</tr>
<tr>
<td>231199</td>
<td>566</td>
<td>33</td>
<td>00:03:51:06</td>
<td>E_FIX 17</td>
</tr>
<tr>
<td>231300</td>
<td>667</td>
<td>00:03:51:09</td>
<td>B_FIX</td>
<td>paper towel dispenser</td>
</tr>
<tr>
<td>231300</td>
<td>667</td>
<td>00:03:51:09</td>
<td>B_FIX</td>
<td>paper towel exit</td>
</tr>
<tr>
<td>231466</td>
<td>833</td>
<td>166</td>
<td>00:03:51:14</td>
<td>E_FIX 21</td>
</tr>
<tr>
<td>231600</td>
<td>967</td>
<td>134</td>
<td>00:03:51:18</td>
<td>paper towel dispenser</td>
</tr>
<tr>
<td>231600</td>
<td>967</td>
<td>00:03:51:18</td>
<td>B_FIX</td>
<td>paper towel exit</td>
</tr>
<tr>
<td>231766</td>
<td>1133</td>
<td>166</td>
<td>00:03:51:23</td>
<td>E_FIX 5</td>
</tr>
<tr>
<td>231800</td>
<td>1167</td>
<td>34</td>
<td>00:03:51:24</td>
<td>paper towel exit</td>
</tr>
<tr>
<td>231800</td>
<td>1167</td>
<td>00:03:51:24</td>
<td>B_FIX</td>
<td>paper towel exit</td>
</tr>
<tr>
<td>231899</td>
<td>1266</td>
<td>99</td>
<td>00:03:51:27</td>
<td>E_FIX 3</td>
</tr>
<tr>
<td>231933</td>
<td>1300</td>
<td>34</td>
<td>00:03:51:28</td>
<td>paper handle slot</td>
</tr>
<tr>
<td>Time</td>
<td>Activity Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>B_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>E_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>left handle right sink</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>B_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>E_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>left hand contact</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>b/w handle faucet</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>right hand under soap disp</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>faucet</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>B_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>E_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>left handle right sink</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>B_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>E_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>left part of soap disp</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>B_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>E_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>right hand under soap disp</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>faucet</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>B_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>E_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>right hand under soap</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td><strong>B_FIX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
- **B_FIX** and **E_FIX** indicate the start and end of activities.
- Times are given in 24-hour format (HH:MM:SS).
237466  6833  0 00:03:57:14  right soap
237466  6833  0 00:03:57:14  B_FIX
237933  7300  467 00:03:57:28  E_FIX
237966  7333  0 00:03:57:29  3
237966  7333  0 00:03:57:29  right hand
237966  7333  0 00:03:57:29  B_FIX
238133  7500  167 00:03:58:04  E_FIX
238266  7633  133 00:03:58:08  17
238266  7633  0 00:03:58:08  left faucet
238266  7633  0 00:03:58:08  B_FIX
238399  7766  133 00:03:58:12  E_FIX
238500  7867  101 00:03:58:15  sink
238500  7867  0 00:03:58:15  B_FIX
238800  8167  300 00:03:58:24  E_FIX
238866  8233  66 00:03:58:26  5
238866  8233  0 00:03:58:26  left faucet
238866  8233  0 00:03:58:26  B_FIX
239000  8367  134 00:03:59:00  E_FIX
239066  8433  66 00:03:59:02  9
239066  8433  0 00:03:59:02  hands
239066  8433  0 00:03:59:02  B_FIX
239933  9300  867 00:03:59:28  E_FIX
240000  9367  67 00:04:00:00  8
240000  9367  67 00:04:00:00  faucet
240000  9367  0 00:04:00:00  B_FIX
240300  9667  300 00:04:00:09  E_FIX
240333  9700  33 00:04:00:10  3
240333  9700  0 00:04:00:10  hands
240333  9700  0 00:04:00:10  B_FIX
244633  14000 4300 00:04:04:19  E_FIX
244699  14066  66 00:04:04:21  13
244699  14066  0 00:04:04:21  left handle right sink
244699  14066  0 00:04:04:21  B_FIX
244766  14133  67 00:04:04:23  E_FIX
244833  14200  67 00:04:04:25  12
244833  14200  0 00:04:04:25  hands
244833  14200  0 00:04:04:25  B_FIX
244566  14833  633 00:04:05:14  E_FIX
245566  14933  100 00:04:05:17  13
245566  14933  0 00:04:05:17  left handle right sink
245566  14933  0 00:04:05:17  B_FIX
245633  15000  67 00:04:05:19  E_FIX
245699  15066  66 00:04:05:21  3
245699  15066  0 00:04:05:21  left handle right sink
245699  15066  0 00:04:05:21  B_FIX
245699  15066  434 00:04:06:04  E_FIX
246133  15500  0 00:04:06:04  33
246133  15500  0 00:04:06:04  left hand contact left handle
246133  15500  333 00:04:06:14  33
246466  15833  0 00:04:06:14  towel dispenser
246466  15833  0 00:04:06:14  B_FIX
246466  15833  0 00:04:06:14  B_FIX
246633  16000  167 00:04:06:19  E_FIX
246733 16100 100 00:04:06:22 handle towel dsp 22
246733 16100 0 00:04:06:22 B_FIX
246733 16100 0 00:04:06:22 E_FIX
247133 16500 12 400 00:04:07:04 B_FIX
247199 16566 0 00:04:07:06 front of towel dsp
247199 16566 0 00:04:07:06 B_FIX
247300 16667 101 00:04:07:09 E_FIX
247300 16667 66 00:04:07:11 tow
247366 16733 0 00:04:07:22 E_FIX
247366 16733 0 00:04:07:22 left hand contact handle tow
247800 17100 20 67 00:04:07:24 front of towel dsp
247800 17100 0 00:04:07:24 B_FIX
247966 17333 166 00:04:07:29 E_FIX
248033 17400 6 67 00:04:08:01 pull down handle
248033 17400 0 00:04:08:01 front of towel dsp
248033 17400 0 00:04:08:01 B_FIX
248233 17600 200 00:04:08:07 E_FIX
248300 17667 11 67 00:04:08:09 exit towel dsp
248300 17667 0 00:04:08:09 B_FIX
249666 19033 1366 00:04:09:20 E_FIX
249733 19100 7 67 00:04:09:22 exit towel dsp
249733 19100 0 00:04:09:22 B_FIX
249933 19200 100 00:04:09:25 E_FIX
249899 19266 6 66 00:04:09:27 towel
249899 19266 0 00:04:09:27 B_FIX
250199 19566 8 300 00:04:10:06 E_FIX
250300 19667 8 101 00:04:10:09 towel
250300 19667 0 00:04:10:09 B_FIX
250300 19667 0 00:04:10:09 E_FIX
250433 19800 12 66 00:04:10:11 towel
250433 19800 8 67 00:04:10:13 B_FIX
250433 19800 0 00:04:10:13 E_FIX
250433 19800 0 00:04:10:13 towel
250933 20300 13 500 00:04:10:28 E_FIX
250933 20300 0 00:04:10:28 rip towel off
251000 20367 13 67 00:04:11:00 towel
251000 20367 0 00:04:11:00 left soap disp
251000 20367 0 00:04:11:00 B_FIX
251399 20766 12 399 00:04:11:12 E_FIX
252199 21566 8 800 00:04:12:06 drying hands
252199 21566 0 00:04:12:06 B_FIX
252733 22100 22 534 00:04:12:22 E_FIX
152

252833 22200 100 00:04:12:25
252833 22200 0 00:04:12:25 left sink
252833 22200 0 00:04:12:25 B_FIX
253000 22367 167 00:04:13:00 E_FIX
253100 22467 100 00:04:13:03
253100 22467 0 00:04:13:03 left wall
253100 22467 0 00:04:13:03 B_FIX
253500 22867 400 00:04:13:15 E_FIX
253566 22933 66 00:04:13:17
253566 22933 0 00:04:13:17 floor
253566 22933 0 00:04:13:17 B_FIX
253699 23066 133 00:04:13:21 E_FIX
253800 23167 101 00:04:13:24
253800 23167 0 00:04:13:24 left wall
253800 23167 0 00:04:13:24 begin turning around
253800 23167 0 00:04:13:24 B_FIX
253933 23300 133 00:04:13:28 E_FIX
253966 23333 33 00:04:13:29
253966 23333 0 00:04:13:29 wall under mirror
253966 23333 0 00:04:13:29 turning around
253966 23333 0 00:04:13:29 B_FIX
254066 23433 100 00:04:14:02 E_FIX
254066 23433 0 00:04:14:02 floor
254466 23833 400 00:04:14:14 turning back toward sink
254466 23833 0 00:04:14:14 B_FIX
254466 24033 200 00:04:14:20 E_FIX
254899 24266 233 00:04:14:27 right sink
254899 24266 0 00:04:14:27 B_FIX
255000 24367 101 00:04:15:00 E_FIX
255033 24400 33 00:04:15:01
255033 24400 0 00:04:15:01 right sink
255033 24400 0 00:04:15:01 B_FIX
255166 24533 133 00:04:15:05 E_FIX
255266 24633 100 00:04:15:08
255266 24633 0 00:04:15:08 front of right sink
255266 24633 0 00:04:15:08 B_FIX
255533 24900 267 00:04:15:16 E_FIX
255633 25000 100 00:04:15:19
255633 25000 0 00:04:15:19 inside sink
255633 25000 0 00:04:15:19 B_FIX
255800 25167 167 00:04:15:24 E_FIX
255899 25266 99 00:04:15:27
255899 25266 0 00:04:15:27 soap dsp
255899 25266 0 00:04:15:27 B_FIX
255966 25333 67 00:04:15:29 E_FIX
256066 25433 100 00:04:16:02
256066 25433 0 00:04:16:02 wall straight ahead
256066 25433 0 00:04:16:02 B_FIX
256200 25567 134 00:04:16:06 E_FIX
256333 25700 133 00:04:16:10 16
towel disp
256333 25700 0 00:04:16:10 B_FIX
256333 25700 0 00:04:16:10 E_FIX
256399 25766 66 00:04:16:12 41
256500 25867 101 00:04:16:15
256500 25867 0 00:04:16:15 right sink
256500 25867 0 00:04:16:15 B_FIX
256600 25967 100 00:04:16:18 E_FIX
256633 26000 33 00:04:16:19 20
256633 26000 0 00:04:16:19 sink
256633 26000 0 00:04:16:19 begin turning around
256633 26000 0 00:04:16:19 B_FIX
256799 26166 166 00:04:16:24 E_FIX
256633 26000 201 00:04:17:00
towel
dsp
256700 26367 0 00:04:17:00 wall/floor
256700 26367 0 00:04:17:00 B_FIX
257166 26533 166 00:04:17:05 E_FIX
257700 27067 534 00:04:17:21 wall just past garbage
257700 27067 0 00:04:17:21 B_FIX
257766 27133 66 00:04:17:23 E_FIX
257833 27200 67 00:04:17:25 8
garbage lid/ego
257833 27200 0 00:04:17:25 B_FIX
257833 27200 0 00:04:17:25 E_FIX
257899 27266 66 00:04:17:27 E_FIX
257933 27300 34 00:04:17:28 3
garbage lid/ego
257933 27300 0 00:04:17:28 B_FIX
257999 27300 0 00:04:17:28 garbage lid/ego
258399 27766 466 00:04:18:12 E_FIX
258433 27800 34 00:04:18:13 4
garbage lid/ego
258433 27800 0 00:04:18:13 B_FIX
258433 27800 0 00:04:18:13 E_FIX
258799 28166 366 00:04:18:24 E_FIX
258866 28233 67 00:04:18:26 12
center garbage lid
258866 28233 0 00:04:18:26 B_FIX
258866 28233 0 00:04:18:26 E_FIX
259299 28666 433 00:04:19:09 E_FIX
259366 28733 67 00:04:19:11 3
towel
259366 28733 0 00:04:19:11 B_FIX
259366 28733 0 00:04:19:11 tow
259633 29000 267 00:04:19:19 E_FIX
259700 567 67 00:04:19:21 8
begin turn
259700 567 0 00:04:19:21 B_FIX
259700 567 0 00:04:19:21 E_FIX
259866 733 166 00:04:19:26 E_FIX
260166 1033 300 00:04:20:05 floor
260166 1033 0 00:04:20:05 B_FIX
260166 1033 0 00:04:20:05 E_FIX
260233 1100 67 00:04:20:07
260366 1233 133 00:04:20:11 15
bottom of door
260366 1233 0 00:04:20:11 B_FIX
260366 1233 0 00:04:20:11
260433 1300 67 00:04:20:13 E_FIX
260466 1333 33 00:04:20:14 14
260466 1333 0 00:04:20:14 bottom left edge of door
260466 1333 0 00:04:20:14 B_FIX
260533 1400 67 00:04:20:16 E_FIX
260600 1467 67 00:04:20:18 15
260600 1467 0 00:04:20:18 middle of door
260600 1467 0 00:04:20:18 B_FIX
260833 1700 233 00:04:20:25 E_FIX
260899 1766 66 00:04:20:27 11
260899 1766 0 00:04:20:27 light switch
260899 1766 0 00:04:20:27 B_FIX
261033 1900 134 00:04:21:01 E_FIX
261066 1933 33 00:04:21:02 8
261066 1933 0 00:04:21:02 wall below light switch
261066 1933 0 00:04:21:02 B_FIX
261166 2033 100 00:04:21:05 E_FIX
261266 2133 100 00:04:21:08 15
261266 2133 0 00:04:21:08 middle of door
261266 2133 0 00:04:21:08 B_FIX
261633 2500 367 00:04:21:19 E_FIX
261700 2567 67 00:04:21:21 4
261700 2567 0 00:04:21:21 door handle
261700 2567 0 00:04:21:21 B_FIX
262066 2933 366 00:04:22:02 E_FIX
262100 2967 34 00:04:22:03 6
262100 2967 0 00:04:22:03 door handle
262100 2967 0 00:04:22:03 B_FIX
262366 3233 266 00:04:22:11 E_FIX
262433 3300 67 00:04:22:13 4
262433 3300 0 00:04:22:13 door handle
262433 3300 0 00:04:22:13 B_FIX
262833 3700 400 00:04:22:25 E_FIX
262899 3766 66 00:04:22:27 10
262899 3766 0 00:04:22:27 middle of door
262899 3766 0 00:04:22:27 B_FIX
263533 4400 634 00:04:23:16 E_FIX
263533 4400 0 00:04:23:16 end
Below is a parsing of the coded data into the Levels that form the hierarchy of tasks for the handwashing task of the same subject (MH).

<table>
<thead>
<tr>
<th>msec</th>
<th>since_start</th>
<th>since_last</th>
<th>timecode</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>230433</td>
<td>-266</td>
<td>-266</td>
<td>00:03:50:13</td>
<td>Start Level 1</td>
</tr>
<tr>
<td>263533</td>
<td>32834</td>
<td>33100</td>
<td>00:04:23:16</td>
<td>End Level 1</td>
</tr>
<tr>
<td>230433</td>
<td>-266</td>
<td>-33100</td>
<td>00:03:50:13</td>
<td>Start Level 2</td>
</tr>
<tr>
<td>230433</td>
<td>-266</td>
<td>0</td>
<td>00:03:50:13</td>
<td>Approach Sink</td>
</tr>
<tr>
<td>232933</td>
<td>2234</td>
<td>2500</td>
<td>00:03:52:28</td>
<td>Turn on Water</td>
</tr>
<tr>
<td>234966</td>
<td>4267</td>
<td>2033</td>
<td>00:03:54:29</td>
<td>Get Soap</td>
</tr>
<tr>
<td>238166</td>
<td>7467</td>
<td>3200</td>
<td>00:03:58:05</td>
<td>Wash Hands</td>
</tr>
<tr>
<td>245466</td>
<td>14767</td>
<td>7300</td>
<td>00:04:05:14</td>
<td>Turn Off Water</td>
</tr>
<tr>
<td>246333</td>
<td>15534</td>
<td>867</td>
<td>00:04:06:10</td>
<td>Get Paper Towel</td>
</tr>
<tr>
<td>250966</td>
<td>20267</td>
<td>4633</td>
<td>00:04:10:29</td>
<td>Dry Hands</td>
</tr>
<tr>
<td>257566</td>
<td>26837</td>
<td>6600</td>
<td>00:04:17:17</td>
<td>Throw Away Paper Towel</td>
</tr>
<tr>
<td>259633</td>
<td>28934</td>
<td>2067</td>
<td>00:04:19:19</td>
<td>Exit Bathroom</td>
</tr>
<tr>
<td>263533</td>
<td>32834</td>
<td>3900</td>
<td>00:04:23:16</td>
<td>End Level 2</td>
</tr>
<tr>
<td>230433</td>
<td>-266</td>
<td>-33100</td>
<td>00:03:50:13</td>
<td>Start Level 3</td>
</tr>
<tr>
<td>230433</td>
<td>-266</td>
<td>0</td>
<td>00:03:50:13</td>
<td>Look at right sink</td>
</tr>
<tr>
<td>233933</td>
<td>3234</td>
<td>3500</td>
<td>00:03:53:28</td>
<td>Left hand contact left handle</td>
</tr>
<tr>
<td>237333</td>
<td>6634</td>
<td>3400</td>
<td>00:03:57:10</td>
<td>Right hand contact soap press bar</td>
</tr>
<tr>
<td>238300</td>
<td>7631</td>
<td>967</td>
<td>00:03:58:09</td>
<td>Put hands together</td>
</tr>
<tr>
<td>240933</td>
<td>10234</td>
<td>2633</td>
<td>00:04:00:28</td>
<td>Hands contact water</td>
</tr>
<tr>
<td>246066</td>
<td>15367</td>
<td>5133</td>
<td>00:04:06:02</td>
<td>Left hand contact left handle</td>
</tr>
<tr>
<td>247666</td>
<td>16967</td>
<td>1600</td>
<td>00:04:07:20</td>
<td>Left hand contact towel pull bar</td>
</tr>
<tr>
<td>250466</td>
<td>19767</td>
<td>2800</td>
<td>00:04:10:14</td>
<td>Left hand rips off paper towel</td>
</tr>
<tr>
<td>251100</td>
<td>20401</td>
<td>634</td>
<td>00:04:11:03</td>
<td>Hands contact paper towel</td>
</tr>
<tr>
<td>259200</td>
<td>28501</td>
<td>8100</td>
<td>00:04:19:06</td>
<td>Left hand contacts garbage front flap</td>
</tr>
<tr>
<td>263000</td>
<td>32301</td>
<td>3800</td>
<td>00:04:23:00</td>
<td>Right hand contact exit door handle</td>
</tr>
<tr>
<td>263533</td>
<td>32834</td>
<td>533</td>
<td>00:04:23:16</td>
<td>End Level 3</td>
</tr>
</tbody>
</table>
References

http://www.arringtonresearch.com


