

Open Loop Pointing in Virtual Environments

by

Barry A. Po

B.Sc. (Computing Science), Queen's University at Kingston, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Computer Science)

We accept this thesis as conforming
to the required standard

The University of British Columbia

October 2002

© Barry A. Po, 2002

Abstract

The two visual systems hypothesis is a powerful neuroanatomical model of the relationship between visual perception and motor action. The hypothesis claims that two independent streams of visual processing within the human brain maintain independent representations of surrounding space. A cognitive stream of visual processing maintains an allocentric map of space, while a sensorimotor stream of visual processing maintains an egocentric map of space. The interactions between these two spatial maps are believed to be responsible for an apparent dissociation between cognitive (non-motor) and sensorimotor responses in numerous artificial settings. Substantial evidence supports the two visual systems hypothesis, ranging from case studies of patients with brain damage to psychophysical studies of normal subjects in tasks that involve the presence of visual illusions.

Because even the most advanced virtual reality systems can unintentionally synthesize visual display artifacts that resemble common visual illusions, the two visual systems hypothesis may be helpful in guiding the design and evaluation of usable interaction techniques for complex virtual reality applications. The hypothesis suggests that inherently cognitive classes of interaction techniques, such as vocal interaction and closed loop (with visual feedback) pointing, are subject to particular execution errors, while inherently motor classes of interaction techniques, such as open loop (no visual feedback) pointing, are not.

This thesis describes an experimental investigation of open loop and closed loop pointing compared to voice-based input in a large-scale interactive display. Experimental results verify that the unintentional illusory errors that are predicted by the two visual systems hypothesis appear with the use of vocal interaction in the presence of a visual illusion known as the Roelofs Effect. Similar response errors were found in closed loop pointing, but there were comparatively fewer illusory errors in open loop pointing. An examination of lag (temporal delay in visual feedback) yielded evidence that lag may also reduce the illusory response errors. These findings support our claim that the two visual systems hypothesis can be influential in improving the usability of virtual reality applications.

Contents

Abstract	ii
Contents	iii
List of Tables	vi
List of Figures	vii
Acknowledgements	viii
1 Introduction	1
1.1 Thesis Contributions	3
1.2 Thesis Overview	5
2 Virtual Reality and the Two Visual Systems Hypothesis	6
2.1 Virtual Reality and Virtual Environments	6
2.1.1 Direct Manipulation and Interaction in Virtual Reality	9
2.1.2 Issues and Limitations with Virtual Reality	13
2.1.3 Motivating a VR Model of Perception and Action	17
2.2 The Two Visual Systems Hypothesis	19
2.2.1 A Short Guided Tour of Visual Perception	19
2.2.2 Ventral and Dorsal Streams: Perceiving “What” and “How”	24
2.2.3 The History of the Two Visual Systems Hypothesis	27

2.2.4	Defining the Two Visual Systems Hypothesis	28
2.2.5	Supporting Evidence for Two Visual Systems	30
2.2.6	Studies with the Induced Roelofs Effect	36
2.3	Perception and Action in Virtual Environments	38
2.3.1	Adopting a Model of Perception and Action in VR	38
2.3.2	The Influence of Visual Illusions and Feedback in VR	42
3	A Two Visual Systems Experiment in Virtual Reality	48
3.1	Goals and Hypotheses of the Experiment	48
3.2	An Experiment in Open Loop Pointing	51
3.2.1	Describing the Virtual Environment	54
3.3	Experimental Presentation	59
3.3.1	Subject Consent and Preparation	63
3.3.2	Cognitive Report Condition with Verbal Responses	64
3.3.3	Motor Report Conditions with Pointing Responses	67
3.3.4	Experiment Session Debriefing and Subject Compensation	70
3.4	Implementing the Experiment	70
3.4.1	Using Audio to Convey Instructions to Subjects	71
3.4.2	Determining the Visual Angle of Visible Objects	73
3.4.3	Using the Polhemus Fastrak as a Virtual Pointing Device	75
3.4.4	Calibrating and Lagging the Fastrak	78
4	Experimental Results	80
4.1	Summarizing the Results	80
4.2	Cognitive Report Condition	84
4.3	Open Loop Pointing Condition	87
4.4	Closed Loop Pointing Condition	90
4.5	Closed Loop Pointing with Interactive Latency	93
4.6	Observations and Subject Comments	96

4.7	Gender and Ordering Effects	99
4.8	Characterizing Subject Motor Performance	101
4.9	An Alternative Analysis of Illusory Effects	104
5	Discussion and Interpretation	110
5.1	Two Visual Systems Phenomena in VR	110
5.2	Developing VR with the Two Visual Systems Model	113
5.3	Why Consider Open Loop Manipulation?	116
6	Conclusion	120
	References	125
	Appendix A Audio Recording Script for Subject Instructions	132
A.1	General Notes and Instructions to Subjects	132
A.2	Cognitive Report Condition Instructions	133
A.3	Open Loop Pointing Condition Instructions	134
A.4	Closed Loop Pointing (Lag and No Lag) Condition Instructions	135
A.5	Cognitive Report Target Presentation Notes	136
A.6	Shared Pointing Target Presentation Notes	136
A.7	Cognitive Report Trial Progression Notes	136
A.8	Open Loop Pointing Trial Progression Notes	137
A.9	Shared Pointing Trial Progression Notes	137
A.10	Condition Concluding Notes	137
A.11	Experiment Session Concluding Notes	138
A.12	Audio Indications and Feedback	138
A.12.1	Verbal Indications to Respond	138
A.12.2	Practice Trial Verbal Feedback	138

List of Tables

2.1	Key Features of the Streams of Visual Perception	29
3.1	Total Number of Trial Presentations in the Experiment	62
4.1	The Induced Roelofs Effect Across Subjects and Conditions	82
4.2	Mean Magnitudes of Subjects with Induced Roelofs Effect	103
4.3	Mean Magnitudes of Subjects without Induced Roelofs Effect	103
4.4	Subject Response Percentages in Cognitive Reporting	106
4.5	Subject Response Percentages in Open Loop Pointing	107
4.6	Subject Response Percentages in Closed Loop Pointing	108
4.7	Subject Response Percentages in Closed Loop Pointing with Lag	109

List of Figures

2.1	A General Overview of the Human Visual System	20
2.2	Retinal Cells in Light Transduction (Dowling & Boycott, 1966) . . .	22
2.3	Ebbinghaus Circles (Top) and the Muller-Lyer Illusion (Bottom) . .	34
2.4	An Example of an Induced Roelofs Effect	37
2.5	A Two Visual Systems Model of Perception and Action in VR . . .	39
3.1	Schematic Overview of the Experimental Virtual Environment . . .	56
3.2	Frame and Target Dimensions for the Presented Visual Stimuli . . .	61
3.3	Determining Visual Angles with Trigonometric Techniques	74
3.4	Virtual Pointing Configuration with Ray Casting	76
4.1	Influence of the Induced Roelofs Effect Across Conditions	83
4.2	Estimated Marginal Means for Individual Cognitive Responses . . .	85
4.3	Scatterplots of Individual Open Loop Pointing Responses	88
4.4	Scatterplots of Open Loop Versus Closed Loop Pointing Variance . .	91
4.5	Comparing the Marginal Means of Both Closed Loop Conditions . .	94
4.6	Comparing the Scatterplots of Both Closed Loop Conditions	95

Acknowledgements

When I started my M.Sc. research some months ago, I quickly came to the realization that writing a thesis at the graduate level is inherently an independent activity with many collaborative elements. There are many people and places that have been integral in getting me to this point, and I would like to make sure that they receive the thanks that they most certainly deserve. I would like to thank my supervisors, Dr. Kellogg Booth and Dr. Brian Fisher, and Dr. Ron Rensink, second reader for my thesis, for providing me with the opportunity, resources, advice, and time that have led me from the beginnings of a thesis research topic to a final product that I consider a personal achievement. Financial support and research facilities for this work were provided by the research and strategic grant programs of the Natural Sciences and Engineering Research Council of Canada (NSERC), by the New Media Innovation Centre of British Columbia (NewMIC), and by the Media and Graphics Interdisciplinary Centre at the University of British Columbia (MAGIC). I would also like to thank Allen Lin for helping me out with all of the smaller intangible aspects of this research, particularly those that involved movement from one place to another. Thanks also go to Alexander Stevenson and Mark Hancock, who have been instrumental in helping me learn everything that there is to know about the Polhemus Fastrak and magnetic trackers in general. Finally, I would like to thank Najwan Stephan-Tozy for her patience and support, which have been central in helping me get this far.

BARRY A. PO

*The University of British Columbia
October 2002*

Chapter 1

Introduction

Our common experience with computer technology reflects the importance of human vision in completing computer-assisted tasks. The very presence of visual information improves our ability to use computers to their fullest potential. In typical desktop computing environments, much of the feedback that we receive about our actions comes in the form of structured pixel data, which are delivered as rendered images that are presented by desktop monitors. Although other perceptual data are undeniably present, particularly auditory and haptic information, we might argue that even in the absence of all other forms of feedback, our sense of vision provides us with the information that is necessary in order for us to determine the effectiveness of our computing actions.

In the context of large scale computer interfaces, such as immersive virtual environments, our reliance on visual sensation increases considerably. Through custom combinations of various direct manipulation devices, virtual environments provide user interfaces that extend far beyond the monitor, keyboard and mouse interfaces that pervade desktop computing. A key feature of many virtual environments is the ability to directly manipulate virtually presented stimuli without having to train ourselves to recall keystroke combinations and without having to browse through complex menu hierarchies. In these kinds of interfaces, we heavily

depend on our sight to guide our movements and actions in a meaningful fashion.

Our motor behaviour and our sense of vision are tied together in a manner that is inextricably linked: our eyes provide input about our surroundings, while our hands, arms, and legs use that input to continually calibrate output movements that are appropriate to the task that we are trying to complete. This modeled relationship between vision and action neatly ties into our naive intuition about visually guided movement. Our intuition leads us to believe that in order to properly orient our movements with respect to objects in our environment, we must first consciously perceive surrounding objects of interest. Conscious awareness of these objects subsequently provides us with the information that we need in order to manipulate them in an appropriate and meaningful fashion.

Our intuition about visually guided motor movement is not entirely accurate. The interaction between our visual and motor systems is significantly more complex than we might expect from casual observations of perception and action. Decades of evidence from experimental psychology suggest a more powerful, neurophysiological model of visually guided motor movement is the *two visual systems hypothesis*, which claims that our conscious awareness of surrounding objects, and our ability to make motor movements toward these objects, are actually separable aspects of visual perception. The two visual systems hypothesis describes how visual information is transduced into signals that are processed by separate regions of the human brain in order to direct our motor behaviour. It also describes the presence of two separate neural streams of visual processing that function to mediate the relationship between visual perception and motor action.

This thesis explores the application of the two visual systems hypothesis to the design and evaluation of complex virtual environments. We examine the influence of visual feedback on pointing interactions in an immersive virtual environment, reporting the results of a user study that involves spatial target acquisition in a display situation where illusory display artifacts are present. The two visual systems

hypothesis predicts that human spatial ability will be compromised under certain conditions in these kinds of environments. The implications of our experimental results are discussed in the context of how we can improve the design and implementation of spatial interaction techniques in future virtual reality applications.

1.1 Thesis Contributions

There are three different areas where this thesis makes substantial contributions. First, this thesis verifies some effects associated with the two visual systems hypothesis in the context of a virtual environment, which is of considerable interest to those involved in human-computer interaction (HCI) or virtual reality (VR) research. This is the new, major contribution of this thesis. Second, this thesis describes novel software support for psychophysical experiments. Third, this thesis demonstrates how a conventional six degree-of-freedom tracking system can be used to construct a prototype gestural pointing interface that is similar to, but is less complicated than, camera-based multidimensional spatial tracking systems described in the literature.

Our experimental evidence strongly suggests that the two visual systems hypothesis has significant implications for the design and evaluation of time and safety critical systems if they depend on direct manipulation or other spatial interaction techniques. The simple awareness of the phenomena that are predicted and explained by the two visual systems hypothesis is the first step in ensuring that preventative measures are put into place to deter undesirable and unexpected user behaviour. This thesis provides sufficient background and evidence to alert designers of virtual reality systems that there are substantial dangers in implicitly assuming simple, intuitive models of human perception and action.

The two visual systems model of perception and action also demonstrates how we can use previously established behavioural theory to guide the development of virtual environments and other user interfaces. The adoption of the two visual

systems hypothesis as a formal model serves as a novel starting point for future VR research that is grounded in established psychological theory. For example, the two visual systems hypothesis suggests that we look into the development of user interfaces that directly introduce perceptual signals into our sensorimotor systems. Such interfaces could have the advantage of further overcoming, or even harnessing to our advantage, the perceptual hurdles inherent in virtual environments. Thus, this thesis provides inspiration to look for other kinds of behavioural phenomena that we might be able to exploit in future interactive designs.

Although user experiments with clear psychophysical underpinnings are fairly common in HCI, the manner in which experiments are administered can benefit from the use of computer technology in order to make the overall experience for participating subjects, and experimenters, more comfortable. Our use of pre-recorded audio instructions shows how a valuable change of presentation can be accomplished through the use of commonly available computer equipment, creating an experimental environment that reduces complexity for the experimenter and reduces fatigue for individual subjects. The presence of details such as these in this thesis motivates others who are conducting similar studies to look for similar ways to improve their own experiments.

Through the use of a Polhemus Fastrak magnetic tracking device as the basis for a gestural pointing interface, this thesis demonstrates how general prototypes that match a wide variety of design profiles can be constructed in a relatively short period of time. Although this is almost certainly not the first time that a VR tracking device has been used in this manner, the implementation details for constructing such interactive interfaces are not widely reported. By including some of the design and implementation details for a Fastrak-based pointing system, future researchers will be able to benefit from the design information that is included here.

1.2 Thesis Overview

This thesis is divided into six chapters. This introduction makes up the first chapter. The second chapter provides background material on virtual reality and the two visual systems hypothesis. A detailed treatment of both VR and the two visual systems hypothesis are provided for those readers who may not have had previous exposure to either, or possibly both, topics. The second chapter also shows how these two topics can be related to improve our understanding of perception and action in virtual environments. The third chapter specifies the design and implementation of a VR experiment that was conducted to verify the existence of phenomena that are classically associated with the two visual systems hypothesis. The fourth chapter presents the observations, results, and subsequent statistical analyses that were derived from the subjects that participated in the VR experiment. The fifth chapter discusses the significance and importance of these experimental results, drawing particular attention to those aspects of the two visual systems hypothesis that are novel to HCI and virtual reality. The sixth and final chapter offers some concluding thoughts.

Chapter 2

Virtual Reality and the Two Visual Systems Hypothesis

An enormous amount of research has been conducted on the topics of virtual reality (VR) and the two visual systems hypothesis. Before we discuss the relationship between these two research topics, we need to cover some relevant background material in each of these two areas. In doing so, we can identify those aspects of virtual reality that are influenced by the two visual systems hypothesis, and we can also see how a model of perception and action that is based on the two visual systems hypothesis could improve the usability of virtual reality systems and advanced user interfaces.

2.1 Virtual Reality and Virtual Environments

Ivan Sutherland (1965) is usually credited with the idea that computer technology would eventually evolve into immersive environments that would provide us with the capacity to interact with computer-generated stimuli in much the same way that we interact with objects in the real world. Over time, his idea of immersive computing led to the emergence of *virtual reality* as a concrete application area of computer science. Although a major portion of contemporary VR systems fall into line with

Sutherland's original vision, the definition of virtual reality has expanded over the years to include all systems that permit us to visualize, manipulate, and interact with sets of data (Aukstakalnis & Blatner, 1992). The idea of a *virtual environment* has emerged alongside VR, indicating that the resulting product of a VR system is an artificial space that permits concurrent observation and interaction with data. Today, computer systems that are based on principles in VR see considerable use in a variety of contexts, including information visualization, real-time simulation, education and training, architecture, medicine, and the life sciences.

Much of the motivation for devising virtual environments comes from our desire to represent complex information in a tangible and accessible fashion. Through the use of VR, we facilitate the presentation of complex data by extending our interactive computing experience to take into account our inherent capacity to experience our surrounding environment in multiple ways. Although the common desktop computer already provides a limited vehicle for the presentation of visual and auditory information, VR systems frequently have a significantly improved ability to access one or more of our internal senses: the tactile, kinesthetic, proprioceptive, and vestibular modalities. By extending the number of ways that computers can interact with us, we simultaneously increase the degrees of freedom with which we may experience information. Ultimately, this means that data that are otherwise too complex to be meaningful can be reduced to manageable form.

The development of VR systems and virtual environments is further supported by a trend in desktop computing toward the inclusion of more VR-like characteristics. Newer graphics cards and processors are capable of rendering complex sets of data to multiple display monitors simultaneously, while display monitors themselves are constantly increasing in size and display resolution. Modern sound cards frequently include some ability to provide sound spatialization, while multi-channel surround speaker configurations are becoming increasingly popular for consumer desktops. Even the input devices of desktop computers are evolving: controllers

with active force feedback motors are more common, and voice recognition is becoming a viable alternative to manual interaction. As a consequence, it is becoming increasingly feasible to suggest that much of the research that currently applies to VR will also apply to desktop environments in the near future.

Complex combinations of display devices and novel interactive tools such as head-mounted and head-coupled stereoscopic displays, data gloves, and other multidimensional spatial tracking devices, are popular examples of VR systems in action. Although such configurations exist and are common in contemporary VR-based systems, virtual environments are not restricted to these stereotypes alone. VR systems range from augmented desktop computing environments, as is the case for fish-tank virtual reality systems, to large-scale room environments, as is the case for CAVE-like arrangements (Ware, Arthur, & Booth, 1993; Browning, Cruz-Neira, Sandin, & DeFanti, 1993). Other more unusual configurations also exist, such as those espoused by the related fields of augmented reality and wearable computing (Starner et al., 1997). With the vast number of VR applications and an ever-increasing number of specialized VR devices, VR systems inhabit a diverse space of nearly infinite configuration possibilities, meaning that there is no single configuration that can be used to describe a “typical” virtual environment.

Nevertheless, every VR system shares a set of common characteristics. Just as our experience with the real world suggests, in order for a virtual environment to function appropriately, at least three different requirements must be fulfilled. First, at least one *output* mechanism must be present, which permits a given VR system to provide us with rendered stimuli that we can observe and experience. Second, at least one *input* mechanism must be present, which permits us to manipulate the stimuli that are presented by the VR system. The presence of these two separate mechanisms implies the presence of a continuous cycle between perception and action: we perceive stimuli through the system’s output mechanism, and we act upon our perception via the system’s input mechanism. Given that the third requirement

is fulfilled, namely, that there is a *consistent relationship* between the system's input and output mechanisms, we have a complete feedback loop, whereby perception continually leads to action and action continually leads to perception.

2.1.1 Direct Manipulation and Interaction in Virtual Reality

The study of spatial motor interaction is an extensive research focus in virtual reality. For many kinds of computing and manipulation tasks, numerous studies have demonstrated that the motor interactivity afforded by virtual environments is significantly better than the keyboard-and-mouse interactivity normally provided by desktop computer systems. When comparing VR-style multidimensional input devices against conventional mouse input devices, experiments by Hinckley, Tullio, Pausch, Proffitt, and Kassel (1997) demonstrated that users were able to complete assigned tasks up to 36 percent faster and without any loss of accuracy when using multidimensional input. Other research by Jacob, Sibert, McFarlane, and Mullen (1994) suggests that user performance improves when the perceptual space of a graphical interaction task mirrors that of the control space of an input device, indicating that the flexibility provided by virtual environments can facilitate the completion of certain manipulation tasks with a level of efficiency that may not be possible through more conventional means.

A major category of motor activities that occur in virtual environments falls under the umbrella of *direct manipulation* techniques. Shneiderman (1982, 1983) introduced the formal paradigm of direct manipulation when he observed an emerging trend toward graphical user interfaces that were mediated by spatial input devices, including pointing implements such as mice, pens, and joysticks. He believed that the emergence of such graphical systems represented a fundamental shift from dialogue-driven interaction to manipulation-driven interaction. Although it is likely that Shneiderman did not have VR systems specifically in mind, it is clear from his definition of direct manipulation that many of the interactive activities supported

by virtual environments are quite applicable to his terminology.

The cornerstone of direct manipulation is the development of a visual language that permits users to manipulate an interactive world. Proponents indicate that there are at least seven benefits that direct manipulation confers: improved learnability, enhanced expert performance, greater memorability, fewer errors, better feedback, reduced anxiety, and increased control (Frohlich, Helander, Landauer, & Prabhu, 1997). There are three key principles behind direct manipulation. First, objects of interest should be continually represented. Second, interaction with objects of interest should be done through physical action instead of complex syntax. Third, the effects of performed physical actions should be immediately visible. Thus, much of the spatial motor interaction in VR systems applies to direct manipulation because of the nature of the input and output mechanisms that make up virtual environments.

However, this does not mean that the interactive techniques used in VR are so easily characterized. The vast amount of research that is dedicated to individual interaction styles indicates that every devised method for interacting with virtual environments has its own set of advantages and disadvantages. Although direct manipulation can be used to group many common VR interaction styles together, it is not entirely clear that the definition of the paradigm alone provides sufficient coverage of the issues that surround the use of one interaction style over another. In addition to recognizing the importance of direct manipulation, we also need to look at a taxonomy of interactive techniques in VR in order to get a good picture of what kinds of features define the basis for spatial motor interaction in virtual environments (Bowman, Johnson, & Hodges, 1999).

Hinckley, Pausch, Goble and Kassell (1994) provide an overview of some of the more prominent issues surrounding the deployment of spatial interaction techniques in virtual environments. Their work focuses on the division of spatial motor interaction issues into two categories: those dealing with human perception, and

those dealing with ergonomic concerns. Most notably, they present evidence to suggest that there is a significant difference between our everyday understanding of interactions in three-dimensional space and our experiencing those same interactions. They find that our capacity to mentally generate solutions to spatial tasks differs significantly from our capacity to physically generate solutions to these same tasks. They identify several different kinds of spatial control metaphors and explore some of the issues that surround dynamic target acquisition in virtual environments.

The work of Poupyrev, Weghorst, Billinghurst, and Ichikawa (1998) builds upon this research to show how most VR manipulation techniques can be segregated into underlying metaphors of either exocentric or egocentric manipulation. *Exocentric* metaphors are characterized by manipulations in relation to global frames of reference while *egocentric* metaphors are characterized by manipulations in relation to ourselves. A key difference between exocentric and egocentric manipulations is that exocentric manipulations occur *outside* of a given environment of interest while egocentric manipulations occur *inside* a given environment of interest. Examples of exocentric interaction include the World-in-Miniature technique, and automatic scaling, both of which involve the external manipulation of objects outside a given environment of interest (Stoakley, Conway, & Pausch, 1995; Mine, Brooks, & Sequin, 1997). Although the research work into exocentric interaction is interesting, the majority of VR interactive techniques involves egocentric manipulation. As such, we focus our attention on these particular metaphors of interaction.

When we refer to the spatial representations of objects in a virtual environment, we may also come across the related concepts of exocentric, allocentric, and egocentric representations of space, which are similar to the definitions described above for manipulation metaphors. In particular, *allocentric* representations of space are similar, but are not exactly the same, as the idea of exocentric representations. While allocentric and exocentric representations both refer to representations of space in global frames of reference, allocentric representations also assume that we

are inside the environment of interest while exocentric representations assume that we are outside the environment of interest. Egocentric representations of space are akin to egocentric manipulation metaphors in that they refer to a spatial representation of environments in relation to ourselves. Thus, when referring to representations of space, we often use the contrasting terms of allocentric and egocentric representations together, unless we explicitly mean that we are representing space outside of an environment of interest.

Egocentric metaphors of virtual environment interaction encompass two basic interaction styles: the virtual hand, and the virtual pointer. Using a virtual hand, we interact with virtual objects of interest by reaching, grabbing, repositioning, and reorienting these objects through a graphical representation of our real hands. In traditional virtual hand interaction, there is a one-to-one correspondence between the movement of our hands and the movement of the virtual hand that is being displayed. A particularly innovative modification of this classic technique is the Go-Go interaction technique, which employs a non-linear mapping in order to extend our volume of reach (Poupyrev, Billingham, Weghorst, & Ichikawa, 1996). Likewise, using a virtual pointer, we can interact with virtual objects of interest by pointing, selecting, and manipulating these objects. Traditional virtual pointing techniques are distinguished from one another by their use of different pointing cursors, varying selection volumes, and methods for determining object selections of interest. Common examples of virtual pointing techniques include ray-casting and virtual flashlights (Bolt, 1980; Jacoby, Ferneau, & Humphries, 1994; Liang, 1994).

Other egocentric interaction techniques fall somewhere in between the virtual hand and the virtual pointer styles. A significant amount of work has been done in developing “natural hand gesture” applications, which improve upon the resolution of virtual hand interaction by permitting users to perform meaningful gestures that include various finger configurations and orientations (Wexelblat, 1995; Ardizzone, Chella, & Pirrone, 2000; Song, Kwak, & Jeong, 2000). Such applications suitably

permit meaningful pointing gestures, but are not limited to these kinds of actions alone. Hand gesturing of this kind has been applied to various VR contexts, including those applications that involve interaction with virtual avatars and navigation through immersive graph data structures (Lee, Ghyme, Park, & Wohn, 1998; Osawa, Asai, & Sugimoto, 2000).

The importance of pointing as an interactive metaphor has been emphasized in recent years. Recent extensions to virtual pointer interaction have come in the form of examining the use of laser pointers as interactive devices for collaboration in virtual environments (Olsen & Nielsen, 2001; Oh & Stuerzlinger, 2002). By combining laser pointers with inexpensive digital cameras such as Webcams, it is now possible to create interactive pointing systems that differ in several key respects from more traditional virtual pointer implementations. In particular, laser pointers are inexpensive, ubiquitous, and wireless. Moreover, their use as interactive devices sidesteps a number of the problems that face other camera-based interaction techniques, such as the need for very high resolution video capture and intensive processing resources. Laser pointers are also highly scalable; some current implementations can handle anywhere from a single user up to an entire room full of users.

2.1.2 Issues and Limitations with Virtual Reality

Unfortunately, virtual reality has yet to reach its fullest potential. This is clearly reflected in the relatively small number of VR systems that are in commercial use. Although many decades have passed since the first virtual environments were conceived, many of the crucial obstacles toward maturity in VR system design still remain. Holloway and Lastra (1993) provide an excellent overview of virtual environment technologies, outlining where VR systems are successful, and where they must still improve. They specifically point out that VR technology suffers from several major problems. Resolution, contrast and dynamic range, field of view, optical

distortion, expense, and operational range are among the many issues that must be resolved with VR display devices before they can become more readily accepted as solid output mechanisms. Spatial tracking and input devices also have a long way to go: accuracy, resolution, environmental interference, effective range, size, robustness, and safety are just some of the areas where VR input mechanisms must be improved.

Such shortcomings are further exaggerated by the fact that the stated design goals of VR devices are often in conflict with one another. For example, increasing the effective field of view for display devices, particularly for head mounted displays, means a corresponding increase in optical distortion and weight. Although displays based on CRT technology might provide relief from some of these issues, high voltages and strong magnetic fields make their use infeasible, particularly when such display devices are used in conjunction with magnetic spatial input devices. Available computing power is also an issue for complex or highly detailed virtual environments. While increases in rendered detail levels, optical fields of view, and stereoscopic imagery add to the experience of presence and immersion, the corresponding computational requirements are also increased, making many VR systems immobile and significantly more expensive.

Interactive Latency

Although a large majority of these technical shortcomings will likely be resolved through improved designs in the future, technological advancement is unlikely to permit VR systems to defy the known laws of physics. Our experience with the real world is unlike our experience with any virtual world because the real world provides us with a seamless correspondence between perception and action. In the real world, the kinesthetic and proprioceptive sensation of our arms is concurrently captured by our sense of vision in the form of actually seeing our arms, and other objects manipulated by our arms, move in space. In virtual environments, such

perfect concurrency is impossible to achieve because of an inherent virtual artifact that is known as *interactive latency*, or *interactive lag*.

Interactive latency exists in virtual environments because of the presence of processing and propagation delays within and between different components of a VR system. The most prominent sources of lag come from the input and output mechanisms of VR devices (MacKenzie & Ware, 1993). Other sources of lag are computational processing time and inter-component propagation. The input and output mechanisms of VR systems contribute to the presence of lag because there are fixed upper bounds on the rates at which input and output devices can be updated. Computational processing time is the amount of hardware or software overhead that is required to translate the information that is received from a spatial input device into a form that can be used to render consistent perceptual feedback. Inter-component propagation is the result of having information pass from one part of a VR system to another. Although these kinds of delays are not very large for uniprocessor and locally-based VR systems, they can be a dominant source of lag for distributed VR systems and collaborative environments that must send information across long-distance network connections.

Individually, each source of lag only produces latencies on the order of several milliseconds. As such, it is highly unlikely that any single source of lag is the reason that we are unable to experience concurrent perception and action in virtual environments. However, when all of these sources of lag are combined, we find that the additive effect is capable of producing delays on the order of one hundred milliseconds, or even higher (Liang, Shaw, & Green, 1991). Experiments into interactive latency by Ware and Balakrishnan (1994) provide compelling evidence to suggest that even such small delays are sufficient to significantly alter our perception and corresponding motor movements in virtual environments.

Only a small number of attempts have been made to accurately model the effects of interactive latency on user motor behaviour in VR systems (MacKenzie

& Ware, 1993). Consequently, only a limited amount of reliable empirical data regarding the effects of lag is readily available. Some of the more prominent studies indicate that display lag causes intersensory discord, leading to errors in both egocentric and exocentric judgments (So & Griffin, 1995a, 1995b). For egocentric judgments, this translates into errors in tracking and following moving targets. For exocentric judgments, this translates into an illusory, apparent motion of a virtual environment (Bajura, Fuchs, & Ohbuchi, 1992).

More common are informal, anecdotal remarks about the ability of lag to cause motion sickness, and its propensity for creating performance-inhibiting environments (Schauffer, Mazuryk, & Schmalstieg, 1996). Even well-known perceptual effects such as oscillopsia, or an inability to visually perceive objects clearly, have been anecdotally reported, although no psychophysical studies have been performed in virtual environments to examine this phenomenon in detail (Bajura et al., 1992). Nevertheless, the informal experiences of first-time VR users generally agree with such assessments; lag is universally accepted as being a negative aspect of virtual reality. Lag is subsequently perceived as a real threat to the usability of systems that employ immersive interfaces, and its presence means that the design of interactive systems that induce a significant amount of lag must either find a way to compensate for its effects, or to convert it in some way to make its presence a positive aspect of the interface.

Even though the effects of lag may not be well-understood in any formal sense, much work has been done to compensate for its effects. Since head-tracked displays are a common feature in VR systems, minimizing and compensating for interactive latency has primarily come in the form of developing models for predictive head-tracking (Liang et al., 1991; Azuma & Bishop, 1994; So & Griffin, 1995b; Kalawsky, 1993; Lawton, Poston, & Serra, 1995; Azuma, 1997). Other attempts to minimize lag have come from research in the development of low-latency spatial tracking systems and low-latency graphical rendering engines (Schauffer et al., 1996;

Olano, Cohen, Mine, & Bishop, 1995). Unfortunately, the effectiveness of these, and any future minimization techniques, are bounded by the fact that we can only minimize end-to-end tracking latency, and not any associated dynamic errors (Azuma & Bishop, 1994).

2.1.3 Motivating a VR Model of Perception and Action

In addition to the technological limitations that must be overcome, there are also barriers of understanding that must be addressed. The majority of VR research has a “design-centric” focus. A significant amount of the effort that has gone into improving our understanding of virtual environments has followed an established process of developing an idea, building a concrete prototype of this idea, and observing the developed prototype’s performance in relation to other prototypes that have already been evaluated. While such an engineering approach has great validity, it also tells us nothing about why a particular prototype has difficulties, nor does it indicate to us how such difficulties can be overcome. Thus, an exploration of VR through a more reflective approach, such as what has emerged in the field of education, and that is encouraged by many practitioners in HCI, may yield answers to questions that cannot be discovered through purely design-based approaches alone (Schon, 1983; Rauterberg, 2000). The integration of a theory-based approach to design in VR allows us to examine how we can take knowledge from other established domains and apply this knowledge to further improve the usability of VR systems.

Our understanding of usability and our realization that user cognition is important to the design of complex interfaces is a reflection of the ever-evolving maturity of HCI research. Developing models of user cognition and behaviour fills an important gap in our understanding of how we relate to virtual environments. To date, there are no widely-reported VR models of user perception and action. It has been suggested that the lack of such models can be attributed to the lack of sufficiently firm theories of human cognition and behaviour (Brown, 1996). Although

this is highly unlikely, given the amount of behavioural theory that exists in the disciplines of experimental psychology and cognitive science, any models that are available are anecdotal at best; they are often based on the intuitive experiences of the designers who initially developed the models. As such, there is no underlying theoretical justification for the validity of the models. Instead, we are simply expected to accept their validity because their definitions happen to fit our intuition about the way that we think we behave.

A perusal of the psychological literature at any depth quickly dispels any belief that our intuition is a reasonable source of understanding about ourselves, or about the environments that we inhabit. Thus, instead of drawing upon our intuition, we might want to consider adopting behavioural models in VR that are grounded in formal, accepted theories of human behaviour. Such models have several advantages over the more anecdotal models that already exist. First, these models are based on a framework of established scientific knowledge, meaning that there already exists good reason to believe that these models reflect true relationships between ourselves and virtual environments. Second, these models are empirically verifiable, meaning that their validity can be inductively supported; this contrasts with more intuitive models, which must rely entirely on subjective rationalization for their acceptability. Third, these models generate testable predictions about our behaviour in virtual environments, meaning that they also suggest the presence of behavioural phenomena that can be exploited to guide and improve the designs of VR systems. Moreover, these predictions permit us to continually evaluate the acceptability of our model. At worst, if a model manages to predict some phenomena and not others, then we at least have a starting point for exploring modifications or alternatives to the model that can capture all phenomena. This is significantly more difficult, if not impossible, in completely design-centric approaches or solely intuitive models.

2.2 The Two Visual Systems Hypothesis

In much the same way that virtual reality has influenced a great number of disciplines both inside and outside of computer science, the two visual systems hypothesis has had a broad impact on many areas of experimental psychology and the behavioural sciences. In evolutionary psychology, the hypothesis has helped us to understand the evolutionary origin of vision, and how these origins have contributed to the development and organization of the human visual system (Milner & Goodale, 1995). It has been equally useful in neurophysiological research because of its ability to help explain the functionality of various regions of mammalian brains. The hypothesis has also been integral in perception research, where it has influenced some of the most fundamental theories about the functions of visual perception. For cognitive science, the two visual systems hypothesis presents a partial framework for understanding the complex relationship between conscious decision making and visually-guided motor behaviour.

2.2.1 A Short Guided Tour of Visual Perception

An understanding of the principles that underlie the two visual systems hypothesis begins with an understanding of the basic components that govern the operation and functionality of the human visual system. Figure 2.1 provides a very general overview of some of the mechanisms that are involved in converting the light information that we receive from our surrounding environment into a form that is appropriate for making decisions and performing actions. Although it is clear that the human visual system is extraordinarily complex, and that our current understanding of its mechanisms is incomplete, many of the fundamental high-level features of visual perception are well-known and are relatively straightforward. In order to understand the basic physiology of visual perception, we necessarily begin with an examination of our eyes, which are the sensory organs of the human visual system.

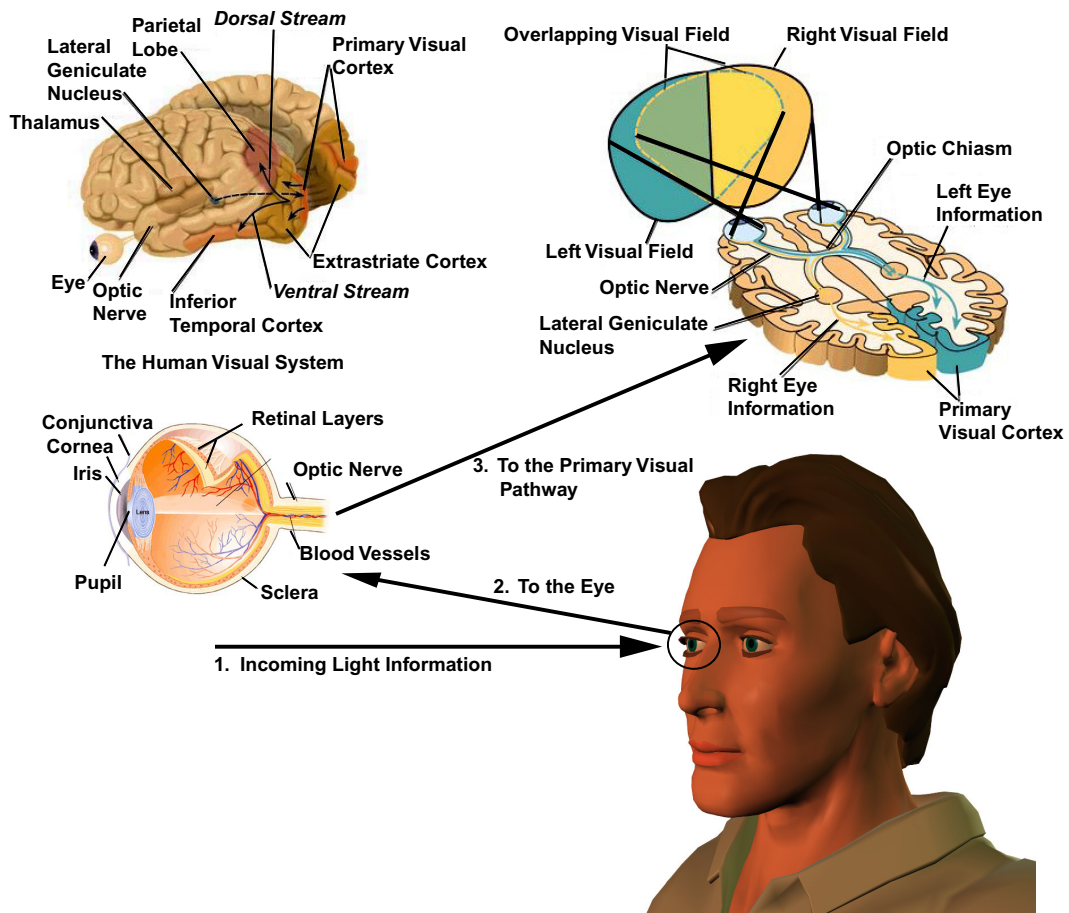


Figure 2.1: A General Overview of the Human Visual System

Despite their enormous physiological complexity, our eyes have a single purpose in the grand process of visual perception: to detect and acquire information about the properties of photons of light energy in our surrounding environment. Our eyes have a horizontal field of view in excess of 200 degrees, and a vertical field of view of approximately 135 degrees (Levine, 1985; Henson, 1993; Goodale & Humphrey, 1998). Our eyes are also extremely acute: under optimal lighting conditions, they are capable of resolving spatial separations of about thirty arc seconds (Bruce & Green, 1990). In order to initiate visual perception, our eyes detect photons of light at different wavelengths and intensities, transducing this detected light

into a form that can be clearly focused at the retina, which is located at the very back of the eye. When the detected light reaches the retina, it is compressed into a form that can be sent down the optic nerve, which is a bundle of nerve endings that connects our eyes to the brain.

At the very front of our eyes are pairs of flexible and fixed lenses that focus the light emitted by objects in our surrounding environment onto the retina. The fixed lens component is more commonly known as the cornea of the eye, and it handles the majority of the light focusing. Through the exertion of torus-shaped ciliary muscles, the flexible lens component of an eye can change shape so that our eyes can adapt in order to focus on objects at varying distances away from us. When the ciliary muscles are relaxed, the flexible lenses are flat, enabling us to focus on distant objects. Likewise, when the ciliary muscles are contracted, the flexible lenses become more rounded, enabling us to focus on close objects.

In addition to having the capacity to change focus, our eyes also have the ability to adapt to varying light intensities in our environment. The iris is the component of the eye that is responsible for this ability. It is a ring that surrounds the pupil, or aperture, through which incoming light must pass before it can reach the retina. Thus, when the incoming light intensity is particularly strong, the iris can constrict the pupil, permitting less light to pass through to the retina. The functionality provided by the iris is especially important because it prevents the retina from becoming saturated by too much light. As a consequence, the iris improves the focus of captured light on the retina. This is a direct result of control over pupil size; when the pupil is smaller in diameter, the area over which light may be scattered over the retina is reduced, leading to captured images with reduced blur.

Once incoming light information has made its way to the retina, three different sets of cell layers are responsible for compressing and transducing these data into action potentials, or signals that activate individual neurons in the brain. Because

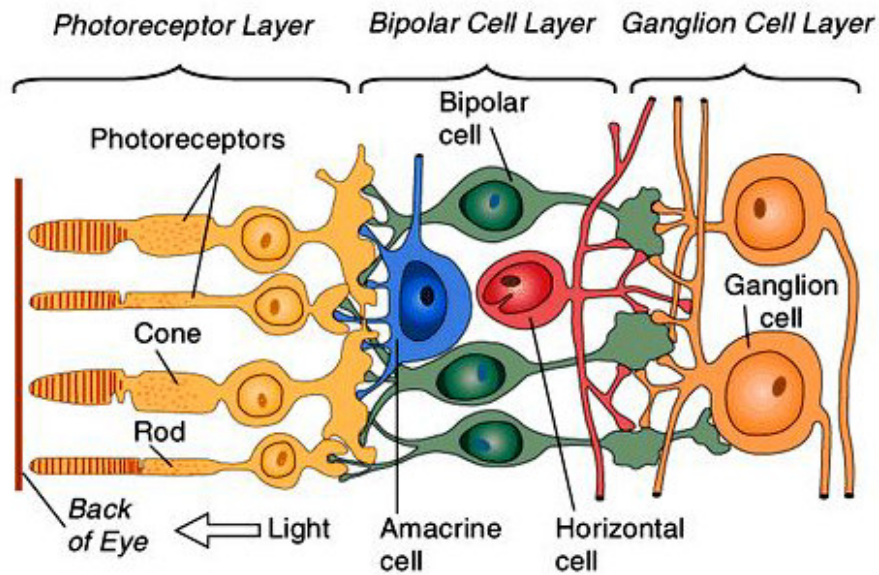


Figure 2.2: Retinal Cells in Light Transduction (Dowling & Boycott, 1966)

this transduction of action potentials is necessary for the activation of the visual processing centres of the brain, the retina is largely responsible for the phenomena that we experience during visual perception and cognition. Figure 2.2 diagrams the cells that are the most important in this process. There are three cell layers of particular interest: a photoreceptor layer, a bipolar layer, and a ganglion layer.

The layer of photoreceptors, made up of rod and cone cells, provides us with sensitivity to light energy of varying wavelengths and intensities. There are anywhere between 75 million to 150 million rod photoreceptors in a human eye, while there are between six to seven million corresponding cone photoreceptors (Riggs, 1971). Rods enable dark-adaptive, scotopic vision that is responsible for the achromatic perception of dimly-lit environments. Cones enable light-sensitive, photopic vision that is responsible for our ability to perceive colour in bright, well-lit environments. With respect to colour perception, there are actually three types of cone cells, each with a different maximum spectral sensitivity: red, or long-wavelength, cones are most sensitive to wavelengths of 575 nanometres, green, or medium-wavelength,

cones are most sensitive to wavelengths of 540 nanometres, and blue, or short-wavelength, cones are most sensitive to wavelengths of 430 nanometres. Together with rod cells, these different kinds of cone cells permit us to detect electromagnetic radiation within a range of approximately 400 nanometres to 700 nanometres.

Light information first reaches the retina through a layer of ganglion cells, which are primarily responsible for receiving the stimuli that are transmitted by the rest of the eye. However, the ganglion cells are also responsible for receiving transduced information from the photoreceptor layer and transmitting appropriate action potentials to the brain. A layer of bipolar cells connects the ganglion cells to the photoreceptor cells that are located at the deepest levels of the eye. Horizontal cells are present in order to define the convergence of signals from particular groups of cones, thereby determining the portion of the visual field that particular ganglion cells transmit to the brain. Amacrine cells are responsible for mediating the same process in rod cells that horizontal cells perform for cone cells: they specify how signals converge for particular groups of rod cells at the ganglion layer, making them very important for visual perception in our periphery.

Once transduced action potentials are generated by the photoreceptor layer and are passed along to the ganglion cell layer, the ganglion cells transmit the received action potentials to two areas of the brain: the lateral geniculate nucleus (LGN) and the superior colliculus (SC). The LGN is located in a region of the brain that is known as the thalamus. It is responsible for organizing information about the properties and locations of objects in our environment and for sending this information to the primary visual cortex. The SC is primarily responsible for projecting information about where objects are in two-dimensional space, allowing us to exhibit the “visual grasp reflex” of turning our eyes and heads toward interesting visual objects (Hess, Burgi, & Bucher, 1946; Ingle, 1973).

The LGN has neurons that project to the primary visual cortex, which is also called the striate cortex or area V1. The primary visual cortex has several

critical functions in visual perception, being the central component of the human visual system. At some point, the images from both eyes must converge to form a single image. When image projections are received from both eyes at area V1, fusion of both images is accomplished, leading to binocular vision, or stereopsis. For VR displays that make use of stereoscopic imagery, area V1 is critical. Additionally, the striate cortex is responsible for separating an image into distinct “feature channels,” where different properties of the environment are encoded as information that is later processed by other higher-level areas of the brain. Specifically, the cortex has cells that extract colour, orientation, and depth information.

The various channels of information that are processed by area V1 are projected to other higher order visual centres for further processing. The primary visual cortex itself has projections to two other centres of visual processing: areas V2 and V3/VP. Area V2 is sometimes called the secondary visual cortex, or the extrastriate cortex. It is responsible for specifically analyzing information about the form and motion of objects in our environment. Because we are interested in the two visual systems hypothesis, area V3/VP is of greater interest to us. At area V3/VP, action potentials that encode visual information diverge to over three dozen areas of the brain, each having some particular function in relation to visual perception. These areas of the brain are frequently identified by two independent streams of visual processing that are differentiated by their general location in the human brain and their category of function. In neuroanatomical terminology, these processing streams are often called the ventral and dorsal streams of visual perception.

2.2.2 Ventral and Dorsal Streams: Perceiving “What” and “How”

Trevarthen (1968), and Ungerleider and Mishkin (1982) are usually credited with presenting the first evidence for a clear neurophysiological distinction between the ventral and dorsal streams of visual perception. Their research identified the two projection streams of vision as broad neuroanatomical features that respectively

encompass the *what* and *where* of visual perception. This distinction between aspects of visual perception suggests that there exists a profound difference in our identification of object shape and form and our identification of object position and orientation. Physiologically, the *what* stream of visual perception was attributed to the functionality of the *ventral stream*, while the *where* stream of visual perception was attributed to the functionality of the *dorsal stream*.

Since the publication of Ungerleider's and Mishkin's work, an explosion of anatomical evidence has confirmed the majority of their claims. The influence of such evidence has been enormous. Because the ventral and dorsal streams show considerable variation in their physiological characteristics, and it has been demonstrated that these variations are responsible for many functional differences, the human visual system is now looked upon as a collection of independent modules, all of which work simultaneously (Goodale, 1994). More recent neuroanatomical studies by Goodale and Milner (1992) suggest that a more appropriate functional distinction between the ventral and dorsal streams is *what* and *how*, as opposed to *what* and *where*; research has also presented credible evidence to suggest that the role of the dorsal stream in perception and action is much larger than originally proposed.

Our current understanding of the ventral stream is that its primary purpose is to provide us with the ability to perceive the shape, size, colour, lightness and location constancies of objects. Thus, the ventral stream contributes all of the necessary components for identifying complex shapes in our environment. The ventral stream maintains an inherently allocentric perception of objects. This suggests that although the ventral stream has a very high acuity for object detail, its ability to spatially reference objects relative to a local origin, such as our body, is quite poor (Jeannerod & Biguer, 1982). Moreover, it is widely believed that the ventral stream is a storage pathway for visual information that must be maintained across different viewing conditions, suggesting a direct connection between visual perception and

spatial memory (Westwood, Heath, & Roy, 2000).

The dorsal stream of visual perception is commonly associated with the movement of the eyes: visual fixation, pursuit and vergence movements, and saccadic eye movements (Busettoni, Masson, & Miles, 1997). Activation of neurons along the dorsal stream has been correlated with the coordination of hand and eye movements and visually-guided reaching movements (Goodale & Milner, 1992). It is believed that the dorsal stream maintains an inherently egocentric perception of objects, and that the dorsal stream is particularly suited toward the maintenance of motor action-relevant information and the processing of the position and orientation characteristics of objects. Because of this, the dorsal stream is thought to serve a mediating role in visual guidance and the integration of prehensile and other skilled motor actions.

Because both the ventral and dorsal streams are simultaneously activated when we engage in tasks that involve visual perception, we experience no apparent segregation of space and form. It has been suggested that the act of spatial attention itself requires involvement from both streams, whether or not that attention is directed toward the task of motor movement, or toward the task of object identification (Rizzolatti, Gentilucci, & Matelli, 1985). Nevertheless, it appears that the ventral stream dominates our conscious visual perception of our surroundings, while the dorsal stream commands an unconscious level of visual perception devoted toward guiding appropriate spatial actions. Some have argued that the dorsal stream never reaches our level of conscious awareness, or at least that it is not required to do so, because such awareness would create interference with the perceptual constancies that are intrinsic to operations within the ventral stream (Goodale & Milner, 1992). Thus, even if we could consciously perceive the dorsal stream in some way, the presence of egocentric information might potentially disrupt the continuity of object identification across changing viewpoints and illumination conditions.

2.2.3 The History of the Two Visual Systems Hypothesis

The original research behind the two visual systems hypothesis did not arise from neuroanatomical studies of human beings. Rather, the hypothesis arose from Trevarthen's (1968) behavioural experiments with split-brain monkeys. His original findings led him to consider the possibility that the vision of object space and object identity in primates were subserved by anatomically distinct brain mechanisms. He called these distinct mechanisms the *focal* and *ambient* processes of the primate brain. The idea of a focal process referred to what is now generally understood as the ventral stream of perception, while the ambient process referred to what is now generally understood as the dorsal stream of perception. Unfortunately, because such knowledge of neurophysiology had not yet sufficiently developed, Trevarthen's terminology was almost universally misunderstood by others to refer to foveal and peripheral vision.

Independent of the work done by Trevarthen, Schneider (1969) came to similar conclusions through his work with brain-lesioned hamsters. Although both researchers provided valuable insight into the physiological basis for vision, their ideas would not be revisited for several years until the more modern neuroanatomical and psychological research of the late 1970s and 1980s. With the work of Ungerleider and Mishkin (1982), the two visual systems hypothesis was allowed to flourish in the form of a supportable, neurophysiological description of the way in which humans process visual information. Bridgeman, Lewis, Heit, and Nagle (1979) sought to more carefully describe the resulting phenomena of separate visual pathways. Their research demonstrated that the distinction between the visual pathways could be characterized as a processing distinction, instead of merely a purely functional one. Specifically, they discovered that meaningful tasks could be assigned to either one of the two streams of visual processing, implying that a distinction did not come from the *kind* of tasks that could be performed by the streams, but rather, by the *responses* that each stream provided for those tasks.

It is clear that the importance of the two visual systems hypothesis to behavioural research has increased over the years. Since its inception, the two visual systems hypothesis has been used to describe the spatial functions of the brain as they apply to our understanding of the way in which internal metrics of spatial information are neurally encoded (Paillard, 1991). In addition, it has been used to help define and validate models of movement within extrapersonal space, or space beyond normal reaching (Jeannerod & Biguer, 1982). The hypothesis has also been used in order to help explain the effects of contextually-influencing visual stimuli in tasks that involve an integration of perception and cognition (Shebilske, 1984).

2.2.4 Defining the Two Visual Systems Hypothesis

With the possible exception of terminology, the basic formulation of the two visual systems hypothesis has seen little alteration over the years. The focal and ambient terms that were proposed by Trevarthen (1968) were superseded in later years by the ventral and dorsal stream distinctions of Milner and Goodale (1992). In experimental psychology and cognitive science, these later terms were recently supplanted by the use of the more domain-appropriate terms of *cognitive* and *sensorimotor* streams of visual perception. Although the neurophysiological aspects of the two visual systems hypothesis are interesting and have added much to the credibility of the hypothesis as a whole, we focus here on the higher-level, cognitive aspects of the hypothesis, which have greater relevance to HCI and virtual reality. For this reason, we will use the cognitive and sensorimotor terminology throughout the rest of our discussion.

A higher-level understanding of the two visual systems hypothesis defines the existence of two functionally distinct representations of the visual world in normal humans (Bridgeman, Peery, & Anand, 1997). A *cognitive* stream of visual perception defines an allocentric map of visual space that is highly sensitive to physical features and properties of objects in our environment. Another *sensorimotor* stream

Perceptual Stream	Function	Duration / Memory Span
Cognitive Stream (Ventral Stream / Focal System)	Conscious visual perception of shape, form, and other physical attributes of objects in the surrounding environment	Continuous (on-going)
Sensorimotor Stream (Dorsal Stream / Ambient System)	Unconscious visual perception of spatial position and orientation attributes of objects in surrounding environment; mediation of visually-guided motor movements	Time-limited (2-8 seconds)

Table 2.1: Key Features of the Streams of Visual Perception

of visual perception defines an egocentric map of visual space that has comparatively poor acuity for physical features; however, it has the capacity to precisely and accurately specify the position and orientation of objects in our environment relative to ourselves, thereby permitting us to make skilled, visually-guided motor movements. We are entirely conscious of the information stored in our cognitive map, but we are unconscious of the information stored in our sensorimotor map. Moreover, the separation of these two pathways is usually opaque to us because these cognitive and sensorimotor streams act in parallel in order to give rise to our conscious perception of and action in the visual world.

The hypothesis claims that there are many instances in which the information that is stored in our cognitive map of visual space differs from the information stored in our sensorimotor map of visual space. In such situations, we lead ourselves to conscious, incorrect conclusions that objects in our environment are positioned and oriented at particular locations that disagree with their actual locations in space. However, if we are asked to act upon this information through the use of visually-guided motor behaviours, the sensorimotor map of visual space leads us to perform correct motor movements, despite our incorrect cognitive awareness. Thus, we may find ourselves in situations where we consciously believe one thing, but unconsciously do something else. Since this phenomenon may be influential in guiding the development of better interaction techniques for complex VR systems, this particular claim is the essence of what we want to study in large scale interactive

environments.

A curious aspect of the separate visual pathways that is only documented in more recent psychological literature is the link between the separate perceptual maps and spatial memory (Bridgeman et al., 1997). It appears that the sensorimotor map of visual space has a time-limited memory, meaning that the sensorimotor representations of objects in our environment may only last for several seconds. Current estimates suggest that this sensorimotor memory of spatial position and orientation may last anywhere between less than two seconds and up to eight seconds, depending on the individual (Milner & Goodale, 1995; Bridgeman et al., 1997). In contrast, it appears that the cognitive map of visual space is not bound by such constraints. Consequently, if we find ourselves unable to retrieve information from the sensorimotor stream that is critical to making appropriate visually-guided motor movements, we automatically rely on the information stored in the less accurate cognitive stream in order to compensate for this lack of sensorimotor guidance.

Table 2.1 summarizes some key features of the cognitive and sensorimotor maps of visual space. More recent research from cortical neurophysiology has determined that there are at least some limited interactions between the cognitive and sensorimotor streams, and that these streams have links to brain areas that are known to be important for human memory (Rossetti & Pisella, 2002). Links have also been discovered between the sensorimotor stream of visual perception and brain areas that are directly responsible for pre-motor cognition and behaviour. These findings are consistent with current beliefs about the nature of the separate visual pathways and are consistent with the two visual systems hypothesis.

2.2.5 Supporting Evidence for Two Visual Systems

Numerous studies and experiments exist to support the claims defined by the two visual systems hypothesis. The majority of this evidence comes in one of two basic forms. First, there are neuroanatomical case studies of disabled individuals whose

perceptive and active faculties are impaired because of brain damage to areas that are critical to one of the cognitive or sensorimotor streams. Second, there are psychophysical experiments of normal individuals, where subjects are exposed to a wide range of visual stimuli that attempt to dissociate the cognitive and sensorimotor streams from one another.

Neurophysiological Evidence: Case Studies

Observing and characterizing impairments that arise in human patients with brain lesions is a common source of useful information in the behavioural sciences. For the two visual systems hypothesis, some of the most provocative evidence comes from case studies of patients who exhibit various inability to perform object recognition tasks or visually-guided motor tasks because of damage to the cognitive or sensorimotor streams. Some of the more prominent studies include examinations of patients who have exhibited symptoms of optic ataxia, visual agnosia, and blindsight.

Optic ataxia is a disorder that is usually associated with damage to an area of the brain that is known as the posterior parietal cortex. The sensorimotor stream leads directly to the posterior parietal cortex, so any damage to this area of the brain should cause a performance deficit in tasks that require visually-guided motor movements. Patients who exhibit symptoms of optic ataxia show just such a deficit; these patients have no difficulty identifying objects or recognizing their physical properties, but they have severe difficulty in directing appropriate reaching movements toward objects. Jeannerod (1986) and Perenin and Vighetto (1988) are credited with reporting some of the first observed instances of optic ataxia. Studies by Jakobson, Archibald, Carey and Goodale (1991) confirmed those observations, describing similar prehensile difficulties in patients of their own. Other cases of optic ataxia continue to be reported in the literature. The most recent evidence suggests that optic ataxia is the result of a disruption of online motor control (Grea et al., 2002; Pisella & Rossetti, 2000), a deficit that fits squarely with the predictions of

the two visual systems hypothesis.

Visual agnosia demonstrates an effect opposite to optic ataxia. Specifically, patients who demonstrate symptoms of visual agnosia have an observed inability to recognize object shape, or any other object physical attributes, though they are quite capable of completing visually-guided motor actions with objects. The most widely cited case of visual agnosia comes from a patient DF, who is reported to have had a bilateral lesion of the occipito-temporal cortex, which is an area of the brain that lies along the cognitive stream of visual perception (Goodale & Milner, 1992). Patient DF was unable to recognize object size, shape, or orientation, and was also unable to successfully complete matching tasks that required her to size her fingers according to the size of visually inspected objects based on a visual representation of these objects. However, patient DF's prehensile motor skill appeared to be quite unaffected by brain damage; when instructed to pick up objects, patient DF was reported to be quite accurate, and her motor movements were observed to be approximately equal in performance to those of a normal individual.

Blindsight is another curious neurophysiological impairment with severe perceptual effects. Patients who exhibit blindsight report a conscious, complete inability to visually perceive objects in their environment. Like patients with visual agnosia, blindsight patients are incapable of recognizing the physical properties of objects. However, unlike patients with visual agnosia, blindsight patients appear to be completely unaware of objects in their environment as well. Blindsight appears to be the result of direct brain damage to the primary visual cortex. Nevertheless, patients diagnosed with blindsight are reported to retain the capacity to orient their eyes and hands to visual stimuli that are briefly presented within the blind portion of their visual field (Weiskrantz, 1986). More recent studies show that blindsight patients can still reach out to supposedly "unseen" visual objects (Perenin & Rossetti, 1996; Jackson, 2000). Because a link between the superior colliculus and the posterior parietal cortex has been used to explain these patients' apparent ability

to perform appropriate motor movements, blindsight is considered to be one more instance of a dissociation between the cognitive and sensorimotor streams of visual perception (Rossetti & Pisella, 2002).

Experimental Evidence: Psychophysical Studies

Experimental studies of normal individuals arguably represent more relevant confirmations of the two visual systems hypothesis in the general population. With studies of normal subjects, we must resort to alternative means for achieving a separation of the cognitive and sensorimotor streams of visual perception. Some of the first studies involving the two visual systems hypothesis came out of experiments built upon a double-step paradigm for studying saccadic eye movements. Later experimental studies evolved toward the use of visual illusions as a basis for creating a dissociative environment for observation.

Studies that are based on a double-step paradigm involve the design of experiments where visual targets are first presented to subjects, and are then displaced during subjects' actions. If the displacements are timed such that they occur in the middle of a saccadic eye movement, then it is expected that subjects will not notice the target displacement (Bridgeman, Hendry, & Stark, 1975). This type of experiment was most often used in studies of saccadic suppression, which examined the phenomenon of apparent perceptual loss during saccades. In one of these studies, Bridgeman et al. (1979) found that subjects were always able to accurately point to the positions of targets, regardless of whether they were able to report a target displacement or not. From this and similar experiments, evidence for two psychophysically separable perceptual streams arose.

These classic experiments later gave rise to another series of studies that explored online arm movement and control. Because it had been suggested that subjects' ability to see their hands during the earlier Bridgeman study could have conceivably influenced results, Pelisson, Prablanc, Goodale and Jeannerod (1986)

replicated the earlier study, with the additional constraint that subjects were presented with targets through an experimental apparatus that hid subjects' hands during the experiment. Even with this change, subjects were still capable of altering their motor movements to compensate for the target displacement. Furthermore, subjects were reported to be unaware that they were making such compensatory movements, in line with their inability to detect the target displacement in the first place.

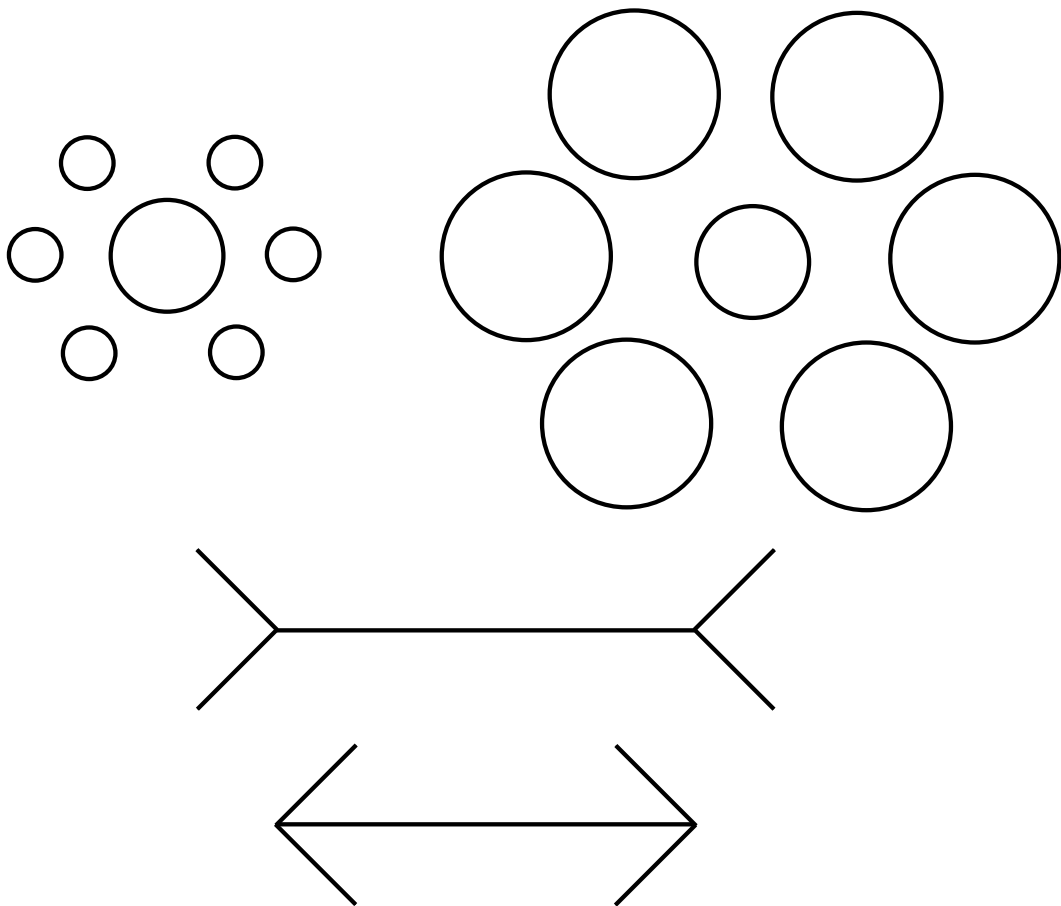


Figure 2.3: Ebbinghaus Circles (Top) and the Muller-Lyer Illusion (Bottom)

Visual illusions provide an interesting basis for the design and implementation of a wide variety of experiments to test the two visual systems hypothesis.

Because such stimuli have the capacity to fool our conscious visual perception, they have found great use as tools for separating the cognitive and sensorimotor streams. Figure 2.3 presents examples of some popular visual illusions. One of the earliest experiments involving visual illusions and the two visual systems hypothesis is a pointing study using induced motion (Bridgeman, Kirch, & Sperling, 1981), which extended earlier findings on saccadic eye movements. For this experiment, a large structured background was continuously displaced during the visual presentation of a small target. Thus, subjects received the impression that the targets that were being presented were moving in a direction opposite to that of the direction in which the background was being displaced. In this particular study, it was discovered that the sensorimotor stream was much less affected by the apparent motion than was the cognitive stream. Thus, it was concluded from the study that apparent target displacement only affected conscious perception, while real target displacement only affected motor behaviour.

Although many of the experiments involving the two visual systems hypothesis are based upon ballistic motor movements in illusory contexts, a substantial number of experiments have also been done involving size assessment and grasping movements. Aglioti, DeSouza, and Goodale (1995) made use of a size contrast illusion called the Ebbinghaus circle illusion, where two circles of the same size are placed in the centre of two circular arrays that are composed of circles of either smaller or larger size. In this configuration, the centre circles appear to differ in size even though they are physically identical. In the Aglioti study, it was observed that subjects would verbally indicate a physical size difference, although during a grasping task of these same circles, subjects' grip size were largely determined by the true size of the circles in question instead of by their illusory size. Other experiments based on different size-contrast illusions, such as the well-known Muller-Lyer illusion where two straight lines of equal length are visually perceived to be of different length because of conflicting depth cues, have reported similarly striking results

(Haffenden & Goodale, 1998; Gentilucci, Chieffi, Daprati, Saetti, & Toni, 1996).

2.2.6 Studies with the Induced Roelofs Effect

Some of the most recent studies into the two visual systems hypothesis have used static visual illusions that are based on a perceptual phenomenon that is known as the *Roelofs Effect*, which is best described as the systematic misperception of target objects that are presented within the decontextualized visual field of an individual (Roelofs, 1935). Visual illusions that are based on the Roelofs Effect enable a dissociation of the cognitive and sensorimotor streams by presenting prominent contextual information about a surrounding environment that biases spatial perception of that environment. Instances of this effect have been reported in several studies outside of research associated with the two visual systems hypothesis; in particular, the effect has been observed as a tendency to perceive the locations of flashing lights as being closer to the line of sight than their actual positions (Mateeff & Gourevich, 1983). By elaborating upon the basis for this effect, it becomes possible to synthesize a visual illusion that is suitable for experiments surrounding the two visual systems hypothesis. These constructed visual illusions are called “induced” Roelofs Effects, which makes a distinction between the evoked illusions and the underlying principle behind their perceptual effects.

Figure 2.4 presents an example of an induced Roelofs Effect, as used in experiments surrounding the two visual systems hypothesis. An induced Roelofs Effect is visually presented in the form of a small, fixed target that is surrounded by a rectangular frame, which is asymmetrically aligned to a given subject. When such a target and asymmetric frame are simultaneously presented, the position of the visual target is misperceived to be closer toward subjects’ midlines than it actually is. Thus, when the presented frame is aligned asymmetrically to the left of a subject, the target is misperceived to be more to the right than it actually is. Likewise, when the presented frame is aligned asymmetrically to the right of a subject, the target

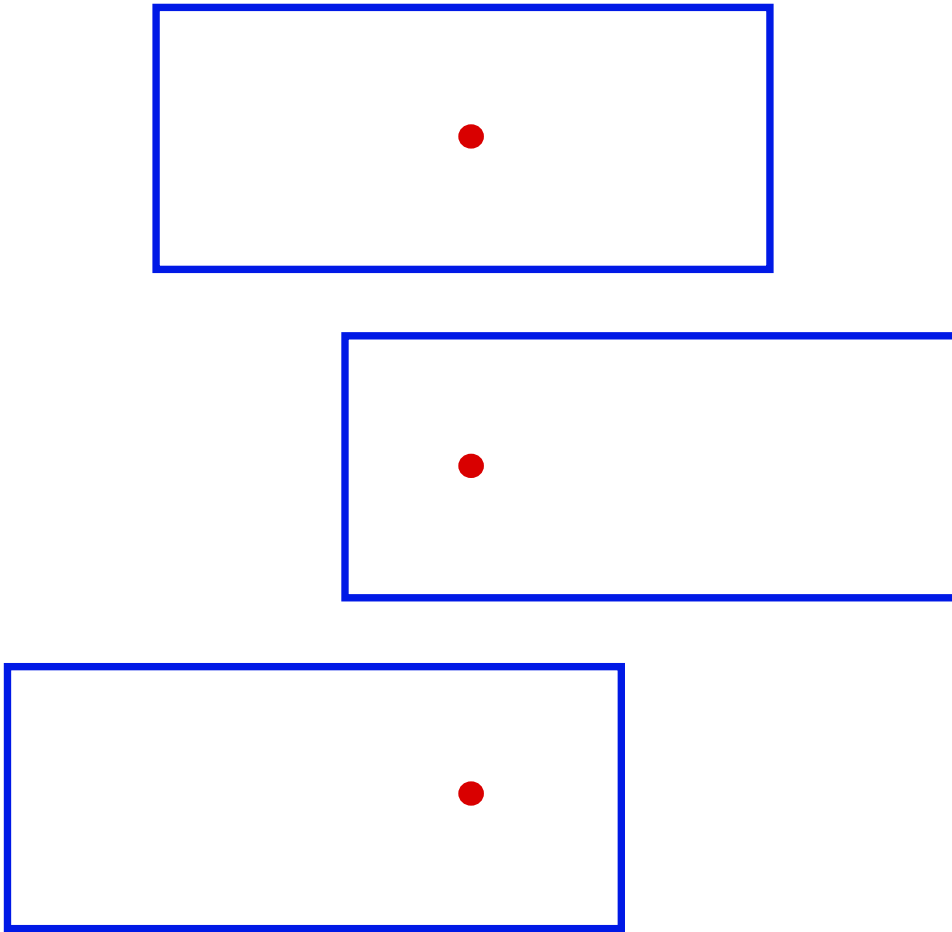


Figure 2.4: An Example of an Induced Roelofs Effect

is misperceived to be more to the left than it actually is. This is the result of making the presented frame a dominant source of contextual information about the surrounding environment. As such, individuals who are likely to rely on visual cues for judgements of spatial position are likely to incorrectly believe that the position of the frame, relative to the position of any presented targets within the frame, provides valid information that can be used to help identify the spatial positions of individual objects in the environment.

The induced Roelofs Effect has been used in more recent studies of the two visual systems hypothesis because it is thought that the presentation of visual il-

lusions that are based on the Roelofs Effect achieves a cleaner separation of the cognitive and sensorimotor streams (Bridgeman, 2000). In addition to this reason, and for at least two other reasons, the induced Roelofs Effect will play a key role in our study of the two visual systems hypothesis and its application to the design and evaluation of complex virtual environments. The first reason is that the induced Roelofs Effect provides the basis for simple, static visual stimuli that are presentable in large-scale display environments. The second reason is that the induced Roelofs Effect creates a reliable perceptual effect that may be generalizable to many tasks that involve direct interactions with complex displays in more applied settings.

2.3 Perception and Action in Virtual Environments

We have seen that the very nature of virtual reality is focused on the interactions of perception and action. We have also seen that the two visual systems hypothesis offers an explanation of the underlying mechanisms that permit perception and action to work in our daily lives. By bringing our knowledge of virtual reality and the two visual systems hypothesis together, it is possible for us to adopt a formal model of perception and action in virtual environments. Such a model is novel because it brings together particular aspects from two distinctly different areas of research. Our ability to rely on the model can be further enhanced because the model can be supported independently by evidence from experimental psychology and virtual reality.

2.3.1 Adopting a Model of Perception and Action in VR

A model of visual perception and action for VR systems necessarily describes the relationship between what we see and how we act in virtual environments. However, in order for such a model to be practical, reliable and useful as well, we need to define it in simple terms with reference only to background material that is directly relevant to the functionality of the model. This means that we should only aim to

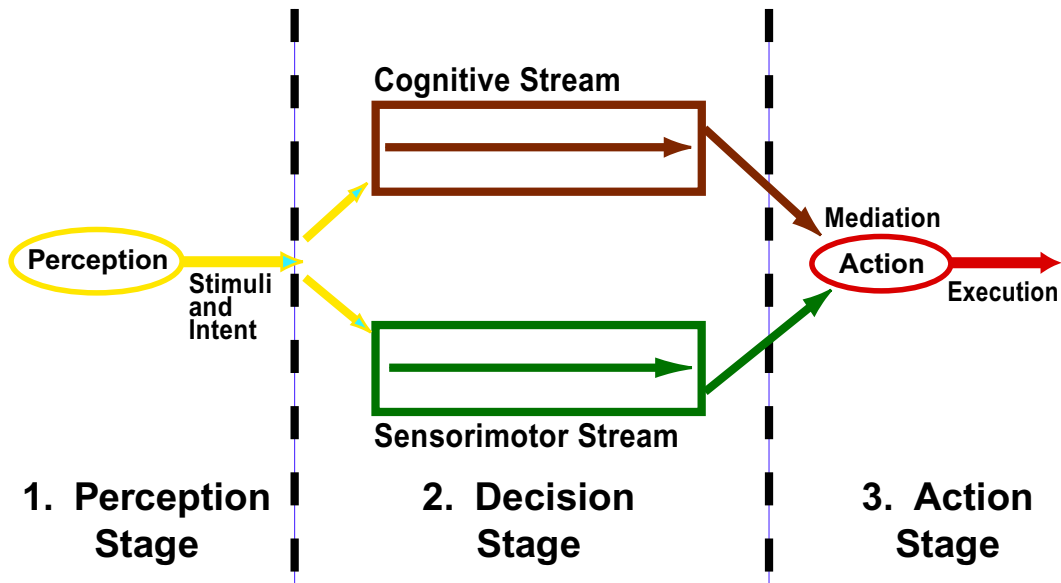


Figure 2.5: A Two Visual Systems Model of Perception and Action in VR

include those relevant higher-level properties of the hypothesis that directly support a description of user cognition and behaviour in VR with the understanding that the underlying justification for these properties exists at lower levels, in the realms of neurophysiology and psychophysics.

Figure 2.5 presents a schematic version of our two visual systems model of perception and action in virtual environments. The model has three successive stages that comprise the overall process of engaging in and completing a single motor action in an arbitrary VR application. The first stage involves the perception of visual stimuli and the registration of intent to perform a motor action in the virtual environment. At this stage, we receive information about our environment and transduce this information into a form that we can use in order to make decisions about appropriate interactions with a VR system. We use the term *intent* in a manner that is consistent with Norman’s (1988) execution-evaluation model in HCI. Although it is debatable from a philosophical standpoint as to what is actually meant by an intention to perform a motor action, we define intent for our purposes to be

any non-reflexive motor action, omitting any formal discussion of its many nuances.

Once visual information is transformed into a form that can be processed, we proceed to the second stage, which describes the decision-making mechanism. This stage is the heart of the model: it describes how we make the transition from perception to action from the perspective of the two visual systems hypothesis. The result of completing the second phase is two output motor configurations from the separate streams of visual perception, each of which specifies a way to perform an intentional motor action in order to interact with a surrounding environment. As we have seen, the cognitive stream is a continually available stream of visual perception that may be used for the performance of visually-guided motor actions. Likewise, the sensorimotor stream is a time-limited stream of visual perception that is also available to make automatic motor configurations for appropriate visually-guided motor actions.

The cognitive and sensorimotor streams provide separate execution pathways that each give rise to possible “decisions” about an appropriate visually-guided movement that should be made in response to our indicated intent. The sensorimotor stream can be thought of as a simple mechanism whose sole purpose is to use primitive visual information about our surroundings in order to formulate an egocentric representation of space that can be used in order to formulate a decision. The cognitive stream can be thought of as a more complex, “evolved” mechanism whose design permits it to perform multiple functions, including the formulation of an allocentric representation of space that can also be used to make motor decisions. However, because of its allocentric design, the cognitive stream is not always well-suited for the task of making decisions about visually guided motor actions. Both streams process and provide answers simultaneously and in parallel. The resulting decisions from both visual streams are passed along to the final action stage, where some mechanism for deciding which of the two answers provided by the streams results in the final action configuration. Both this mechanism for mediation of the

two visual streams and the inner workings of both visual streams are intentionally abstracted in our model in order to avoid concerning ourselves with the lower-level workings of the mediation process. While these details are undoubtedly interesting, we remain most concerned here with the model's ability to outline the relationship between perception and action because it does not matter to us how these particular processes are accomplished.

We must be careful to draw a distinction between the cognitive stream and the complex entity that defines human cognition. This distinction is not often made in research on the two visual systems hypothesis and this omission has been used as an argument against the prescribed definition of the cognitive stream. Proponents of such contrary arguments suggest that the use of the term "cognitive" may actually be inappropriate (McGeorge, 1999; Jeannerod, 1999). Thus, our definition of the cognitive stream relates only to its prescribed function of conscious perception of environmental attributes and its allocentric representation of space in the two visual systems hypothesis. While we would be remiss to suggest that true cognition plays only this limited role in the determination of final motor movements, an elaborate discussion of the possible roles and influences that human cognition may have in determining visually guided motor behaviour is definitely beyond the scope of this thesis. It is only relevant here that cognition may be an influential component in determining appropriate motor action.

A critical property of our model is that there is no implicit assumption that our conscious beliefs about motor actions being taken and the motor actions actually being performed have any correspondence. This bears some similarity to Norman's (1988) distinct notions of *errors of intent* and *errors of execution*. We recognize that there is a distinct difference between what we may intend to do, and what we actually do. Thus, the resulting decisions generated by our model take into account the characteristic property of the two visual systems hypothesis that both the cognitive and sensorimotor streams of visual perception may simultaneously hold different

perceptual views of the real or virtual environment. This is particularly important because it reinforces the belief that our intuitive understanding of perception and action in VR is incorrect. Furthermore, it provides some reason for seriously considering a more formal model: an intuitive model of perception and action would not have been able to predict such a potentially influential phenomenon, nor would it even suggest that such a phenomenon existed in the first place.

2.3.2 The Influence of Visual Illusions and Feedback in VR

It may not be immediately clear how the two visual systems hypothesis and our two visual systems model of perception and action can inform the design and evaluation of complex virtual environments. However, if we take a closer look at how certain interaction techniques and commonly followed interface conventions in VR applications are influenced by the perceptual and cognitive factors that are described by the two visual systems hypothesis, it becomes clear that our model of perception and action has a lot to offer. Of particular interest are those methods of interaction that use inherently cognitive and inherently motor response mechanisms to interact with virtual environments. We briefly list a few of these interaction styles here.

Assuming that the speed and accuracy of current voice recognition technology continues to improve, Oviatt and Cohen (Oviatt & Cohen, 2002), among other supporters of multimodal user interfaces, suggest that vocal interaction may be an efficient and accessible technique for interacting with complex systems. Based on several previous studies of the two visual systems hypothesis, vocal interaction is recognized as an inherently cognitive response mechanism, forcing users to utilize their cognitive spatial maps in order to make judgments about spatial relationships between objects (Bridgeman et al., 1997; Bridgeman, 2000). As such, the two visual systems hypothesis suggests that interactions involving the vocal localization of target objects in an interactive display environment are subject to systematic errors in user judgment because users' cognitive maps of space can be biased by the

perceptual effects of visual illusions.

These same studies also suggest that motor interaction techniques, such as pointing without visual feedback of any kind, rely entirely on the spatial map that is provided by the sensorimotor stream of visual processing, and not at all on the spatial map provided by the cognitive stream. In the presence of visual feedback, the two visual systems hypothesis predicts that there will be increased interactions between the cognitive and sensorimotor maps of space, which may lead to unnecessary errors of execution that are brought upon by the apparent dominance of the cognitive representation of space. The two visual systems hypothesis provides a suggestion that the presence of visual feedback in user interaction with virtual environments is influential in determining the overall robustness of the interaction techniques that we choose to use.

For these and other reasons, our model of visual perception and action presents some motivation for examining particular characteristics of virtual environments that may directly influence our behaviour at an unconscious level. Our model suggests that we should more closely examine the kinds of visual feedback that we present in response to our motor actions, and that we should also be wary of any intuitive belief that visual feedback mechanisms are always helpful to us when performing visually-guided motor actions. For example, our model indicates that we should be aware of the dangers of introducing unintentional visual illusions or other visible artifacts into our visual field while we are working in a virtual environment. In this case, the model suggests that any perceptual inconsistencies introduced by VR systems may directly affect our ability to consciously recognize the difference between what we see and how we actually interact with our environment.

Effects of Visual Artifacts

Recall that psychophysical studies of normal human subjects involving the two visual systems hypothesis frequently make use of visual illusions in order to dissociate

the cognitive and sensorimotor streams of visual perception from one another. Such illusions are used because dissociation of the visual streams is difficult to observe and analyze in real world environments. Most likely, in a real world environment, the effects of visual illusions are rarely observed because our sensory experience with the real world rarely entails interaction with such directed visual stimuli, unless such stimuli are deliberately presented, as they are in controlled experiments. Visual illusions are likely made inseparable from other visual stimuli in our real world experience because of the enormous amount of perceptual information that we are able to derive from various details in our surroundings. Thus, in a real world environment, the presence of two segregated visual pathways may not impact to any great degree our ability to perform many daily motor tasks.

However, there are reasons to believe that visual illusions are a more frequent occurrence inside virtual environments and that contextual information is very important to our understanding of space in immersive environments. Context has been shown to induce differences in perception, most notably in visual illusions such as the moon illusion (Shebilske, 1984). Complex graphical renderings of scenes or other complex sets of data are often accomplished through the presentation of points, lines, and simple polygons. Despite the display resolutions and levels of detail at which such graphics are now being rendered, current VR systems simply cannot match the richness of real world environments. As such, certain virtual presentations of scenes may inadvertently exhibit characteristics of visual illusions that may not be present in an equivalent real world environment. When these are combined with the potentially different interactions that we have with virtual environments that have no parallel with the way in which we interact with the real world, we begin to see how an awareness of the influence of visual illusions must be more important.

For example, as seen at the bottom of Figure 2.3, the perceptual effects of the Muller-Lyer illusion are commonly attributed to our visual experience with continuous straight edges and linear perspective cues of depth perception (Coren

& Girgus, 1978). In architectural applications that may involve the manipulation, perhaps by interactions such as dragging or scaling through a virtual hand metaphor, of building walls or other similar structures that have been rendered with a typical perspective projection, it is conceivable that we might be negatively influenced by the perceptual effects of the Muller-Lyer illusion. In this particular instance, the results of illusory influence may exhibit themselves in the form of inadvertently resizing building structural supports because we incorrectly perceive them to be unequal in height. In planning and design applications, the consequences of such misguided actions might mean unnecessary time being spent in correcting these mistakes, and in the worst case, it may even mean that unsafe buildings are put forward for construction. Such an example is only one demonstration of how visual illusions could impact user cognition and behaviour inside a virtual environment.

Additionally, many VR applications are designed around interfaces that do not attempt to mimic the real world, but instead, attempt to immerse users in a functional environment that is specific to a particular task. Many applications, including geographic information system applications, collaborative environments, and educational learning systems use such an approach. These kinds of applications provide a wealth of opportunities for undesirable visual illusions that are recognized as having different cognitive and sensorimotor effects. In these kinds of virtual environments, tasks such as target acquisition and manipulation are very common, showing many common characteristics with past psychophysical studies of the two visual systems hypothesis and target acquisition. Therefore, it is conceivable that in such applications, visual stimuli that exhibit illusory characteristics may cause dissociations of the cognitive and sensorimotor pathways. With this motivation, we have investigated a specific question about whether or not our conscious perception of what we do can differ from what we actually do during specific interactions for localizing targets on a large scale display.

Open Loop and Closed Loop Virtual Environments

Visual feedback comes in many forms in virtual environments, although the specific type and amount that are provided vary from application to application. One of the most prominent uses of visual feedback is to provide guidance for ballistic aiming movements and for target acquisition tasks. The virtual hand and virtual pointer interaction metaphors typically provide an aiming mechanism in the form of a rendered hand, a targeting crosshair, or some other iconic guide in order to support application-specific tasks. Strictly speaking, visual feedback is also provided in response to an executed motor movement that alters a virtual environment in some fashion, meaning that we can see the results of our actions. We also see visual feedback in the form of our arm and hand movements when we make reaching or pointing motions. Other body movements present similar kinds of visual feedback during the course of other types of motor interactions.

The presence of visual feedback for motor action in a virtual environment defines an additional feedback loop between user perception and action. When this visual feedback is provided in a VR system, the resulting virtual environment is said to be *closed loop*. This refers to the fact that visual feedback, regardless of form, facilitates a continuous corrective mechanism that permits users to make fine adjustments to corresponding motor interactions according to visual perception in the virtual environment. When such visual feedback is not present, the lack of visual feedback breaks the feedback loop between user perception and action. Virtual environments of this type are said to be *open loop*. When corresponding motor actions are said to be accomplished with or without aid of visual feedback, those motor actions are called closed loop and open loop motor actions respectively.

Strictly speaking, the lack of visual feedback does not create a purely “open loop” virtual environment. Perceptual information is still available in the form of internal sensations, such as tactile feedback, kinesthetic sense, and proprioception. Although the two visual systems hypothesis and our adopted model for VR do not

specifically address these other forms of motor feedback, it is important to recognize that these internal senses are available to create an internal feedback loop that permits some level of motor correction. Because we are focused on the relationship between *visual* perception and motor action, our definitions of open loop and closed loop virtual environments are similarly focused such that we avoid referring to these other sensory channels in our definition even though they likely play an important role in sensorimotor behaviour.

Intuition appears to favour closed loop virtual environments much of the time, particularly when pointing interactions and target acquisition tasks are concerned. Most people have a lack of confidence in completing such tasks without visual feedback; there may be a feeling that attempting to accomplish such tasks without feedback is the equivalent of attempting to accomplish them while effectively blind. Past experimental evidence from the two visual systems hypothesis suggests otherwise: because open loop motor performance does not appear to be affected by the perceptual effects of visual illusions, this form of interaction may actually be more accurate, though not necessarily more precise, than closed loop motor performance under certain conditions. The lack of visual feedback prevents interference from the cognitive stream, whose influence may improve the precision of particular motor movements, although its allocentric nature may affect overall accuracy. In terms of direct manipulation in VR, this suggests that an examination of the characteristics of open loop versus closed loop interactions can teach us how to predict certain performance characteristics that underlie individual spatial interaction techniques in virtual environments. In the next chapter, we describe a particular experimental design to test one such prediction.

Chapter 3

A Two Visual Systems Experiment in Virtual Reality

Virtual reality introduces us to a wide range of different visual experiences that may create dissociations between visual perception and motor action. If we can demonstrate that the classic effects that are seen in past psychological experiments under strict experimental conditions can also be demonstrated in virtual environments that are significantly richer in visual context, then we also provide good evidence to support the claim that known implications of the two visual systems hypothesis also apply to the development of VR systems. Interpreting user performance in virtual environments in the context of the two visual systems hypothesis can augment our understanding of user behaviour, and an exploration of open loop motor performance may lead to potentially significant changes in the way that we think about direct manipulation in VR.

3.1 Goals and Hypotheses of the Experiment

Establishing criteria for validating our two visual systems model of perception and action in virtual environments is a matter of looking at what goals were established in past experiments of the two visual systems hypothesis and adapting those goals to

the task of informing the design and evaluation of complex user interfaces. Drawing from past studies in experimental psychology, we observe that all of the concrete experimental evidence for the two visual systems hypothesis in normal subjects comes from being able to demonstrate that there is a dissociation between the cognitive and sensorimotor streams of visual perception. Such a dissociation can be expressed as subjects exhibiting an incongruency between what they perceive and how they act on what they perceive.

Dissociations of this kind can be more formally expressed in two distinct experimental hypotheses. First, we need to demonstrate that *subjects are capable of consciously misjudging the locations of presented objects in virtual environments*. With this hypothesis, we expect to test for the presence of functionality that is provided by the cognitive stream of visual perception. Through the use of a visual illusion that is known to bias the perceived location of a small target object, and by using a verbal mechanism for reporting perceptual experience, confirmation of this first hypothesis is possible. If we manage to demonstrate the correctness of this particular hypothesis, we also demonstrate that our visual awareness of a virtual environment can be subject to perceptual errors.

Second, we need to demonstrate that *subjects are capable of accurately performing motor actions toward these same presented objects through open loop motor behaviour*. With such a hypothesis, we expect to test for the presence of functionality that is provided by the sensorimotor stream of visual perception. Since previous psychological studies indicate that the presence of visual feedback may create interference between the two streams of visual perception, it is vital that this condition be carried out under open loop conditions in order to ensure that we are testing a true dissociation of the cognitive and sensorimotor stream (Pelisson et al., 1986). By using the same visual illusion in conjunction with the same target task as in the first hypothesis, but with a specific motor mechanism for reporting sensorimotor behaviour, confirmation of this second hypothesis is possible. If we manage to

demonstrate the truth of this second hypothesis, we also show that our ability to perform motor actions in virtual environments is unaffected by biases introduced by visual illusions, despite the presence of conscious perceptual errors.

An examination of open loop motor performance is also important for more pragmatic reasons: it may tell us whether manipulation without visual feedback is feasible, and whether such a form of manipulation should be actively pursued in the development of certain VR applications. Thus, through determining the validity of these two hypotheses, we are testing for a dissociation that replicates effects that are reported in past experiments of the two visual systems hypothesis, and we are demonstrating that there are results that stem from the two visual systems hypothesis that may prove useful in the design and evaluation of robust spatial interaction techniques. These two experimental hypotheses are tested under conditions that are representative of large scale, immersive environments, so we also provide evidence that supports applications of the two visual systems hypothesis in developing VR applications.

Although these two hypotheses provide sufficient criteria to judge the validity of a two visual systems model of perception and action in VR, we are also interested in seeing how we might be able to interpret the two visual systems hypothesis so that we can apply it more readily to the development of VR applications. Provided that we have evidence that shows observable effects that can be attributed to the two visual systems hypothesis, we can further examine the applicability of the hypothesis by testing additional experimental hypotheses. In particular, a two visual systems model of perception and action suggests that *open loop motor performance is more accurate than closed loop motor performance in certain kinds of VR applications*. This comes from an understanding that the cognitive stream of visual perception interferes with the resulting motor outputs of the sensorimotor stream by inducing unnecessary motor corrections to visually guided motor actions. If this experimental hypothesis is true, then the two visual systems hypothesis provides a useful expla-

nation of this phenomenon. By testing a closed loop form of the manipulation that is used to test open loop motor performance and searching for evidence of motor errors that are akin to observed perceptual errors, we can determine whether closed loop motor performance is subject to errors that may not be present under open loop conditions.

A further prediction of the two visual systems model is that because VR systems with visual feedback motivate closed loop motor performance, we may be able to detect behavioural phenomena that are associated with the presence of interactive latency. Specifically, a two visual systems model can be used to predict that *the presence of lagged visual feedback may induce subjects to perform similarly to open loop conditions*, despite the presence of visual feedback. This comes from our understanding that the influence of the cognitive stream of visual perception is particularly dependent on reliable visual feedback, so that the introduction of unreliable visual feedback may significantly reduce this reliance on the cognitive stream and revert to the sensorimotor stream of visual perception instead. If we see evidence to suggest that manipulation with lagged visual feedback demonstrates a lack of motor error when compared to any potential motor errors in the open loop and closed loop conditions, we can determine the validity of this particular experimental hypothesis.

3.2 An Experiment in Open Loop Pointing

In order to test the outlined hypotheses, we used a within-subjects behavioural experiment with a psychophysical framework. A within-subjects design was used because of its increased sensitivity to effects that were present in the psychophysical tasks that subjects were expected to perform. Additionally, such an experimental design conferred specific analytical advantages that could not be achieved with a between-subjects design. A within-subjects experiment provided us with the capacity to gauge the performance of each subject across all of the different interaction

styles, and it allowed us to look at these performance differences across all subjects, which is particularly significant for HCI research because it gave us a way to comprehensively assess the advantages and disadvantages of individual interaction techniques under different conditions.

Fourteen subjects were elicited to participate in the collection of results and data. With one exception, all of the subjects were undergraduate students, graduate students, or recent graduates from various faculties and departments within the University of British Columbia. Of these participants, seven subjects were male and seven subjects were female. All of the subjects were between the ages of 17 and 31. The subjects had varying levels of experience with VR-type applications. Twelve of the subjects claimed to be right-hand dominant, while two of them claimed to be left-hand dominant. All of the subjects had normal, or corrected-to-normal vision in the form of glasses or contact lenses. Five subjects reported normal vision, six subjects wore glasses, and three subjects wore contact lenses during the experiment. Across all of the subjects, the duration of the experiment sessions ranged from approximately one hour and thirty minutes to two hours.

A first step to verifying the two visual systems hypothesis in VR is a simple replication of past psychological studies on the two visual systems hypothesis. However, it is important to realize that because these past studies were only interested in the behavioural effects of the hypothesis and were not concerned at all with general applicability, all of these past experiments were done in relatively small, enclosed spaces with extreme control over ambient lighting and other environmental properties that were attuned toward maximizing the effects of cognitive and sensorimotor dissociation (Aglioti et al., 1995; Bridgeman et al., 1997). In effect, these past experiments are not necessarily reflective of the kinds of situations that arise in virtual reality. Because we were interested in observing the power of our model through additional hypotheses that have no parallel in past studies into the two visual systems hypothesis, we needed to design a new experiment that took place within a

virtual environment, drawing upon the experiences of past studies wherever it was necessary and reasonable to do so. In particular, we needed to differentiate our experiment from previous work by situating participating subjects in an environment that was clearly configured for virtual reality applications and by providing subjects with tasks that were representative of those that are used in VR applications.

Our present experiment was organized around four distinct conditions, each of which examined the behaviour and performance of subjects in a different interactive situation. The basic task that underlied all of these conditions was the acquisition of visually-presented targets in a virtual environment. Target acquisition was used because it represented an archetypical direct manipulation task that is common within many VR applications. From a psychological standpoint, target acquisition is a task that can be performed through inherently cognitive or inherently sensorimotor response mechanisms. From a VR standpoint, vocal localization and pointing are inherently cognitive and inherently sensorimotor interaction styles that can be used to accomplish target acquisition and manipulation tasks. Thus, it appeared quite reasonable to assess the validity and power of our two visual systems model through just such a task.

One condition involved a verbal report of several distinct target positions and was dedicated to testing the hypothesis that we are capable of misperceiving object positions in a virtual environment. For this reason, we called this particular condition the *cognitive report* condition. A second condition involved open loop motor report of the same distinct target positions and was dedicated to testing the hypothesis that conscious misperception of object positions has no equivalent perceptual effect on actual motor performance. In this condition, subjects were asked to make ballistic pointing motions in order to acquire targets, effectively substituting the verbal report of the cognitive report condition with a form of direct motor report. We called this particular condition the *open loop pointing* condition. A third condition modified the open loop motor condition, providing visual feedback in the form of a

crosshair that corresponded to subjects' pointing movements, testing the prediction that open loop motor performance is more reliable than closed loop motor performance. This particular condition was called the *closed loop pointing* condition. The open loop pointing and closed loop pointing conditions were counterbalanced such that half of the subjects experienced the open loop pointing condition before the closed loop pointing condition, while the other half of the subjects experienced the closed loop pointing condition before the open loop pointing condition. This was done to avoid possible ordering effects between the two conditions. A fourth condition made an additional modification to the closed loop motor condition, adding a measure of interactive latency to the crosshair that was provided as visual feedback. As a result, this condition tested the prediction that lagged visual feedback may induce open loop motor performance. This condition was called the *closed loop pointing with interactive latency* condition.

For our experiment, the induced Roelofs Effect was selected to provide the visual stimuli that were necessary for our experiment. Because the induced Roelofs Effect was successfully used in the most recent studies of the two visual systems hypothesis in order to demonstrate dissociations of the cognitive and sensorimotor streams of visual perception, we had good reason to believe that such an illusion would perform the same function in virtual environments. Because this particular illusion also provided a convenient visual presentation that was easily adapted for use in pointing and target acquisition tasks, it was a reasonable choice for stimulus presentation in our experiment.

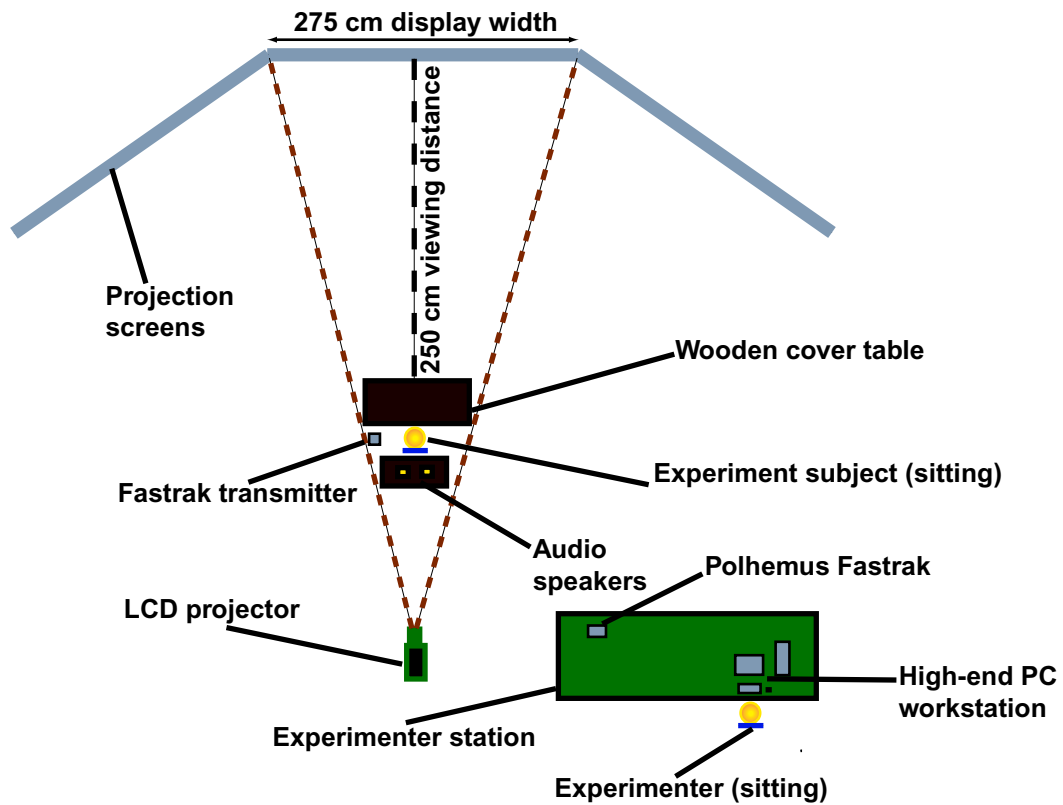
3.2.1 Describing the Virtual Environment

Since our goal was to apply the two visual systems hypothesis to virtual environments, choosing an appropriate experimental apparatus and environment configuration was critical to our experimental design. The design did not explicitly call for specific VR technology, meaning that any choice, ranging from the scale of a wear-

able, head-mounted environment to that of a large scale, immersive environment might have been sufficient. However, because a significant number of novel egocentric target acquisition tasks occur in large scale VR systems, and because large scale systems frequently provide immersion in a relatively decontextualized environment with few external perceptual distractors, we chose to use a large scale, immersive environment as the setting for this experiment.

Figure 3.1 provides a schematic overview of the virtual environment that was used for our experiment. Our apparatus was set up in the Landscape Immersion Laboratory, located in the Forest Sciences Centre at the University of British Columbia. This lab provided a suitable location for constructing an appropriate immersive virtual environment configuration. The physical arrangement of the lab includes an open display facility that is suitable for high-end multimedia presentations, landscape simulation, architectural design, and immersive virtual reality applications. The centrepiece of the lab is a set of three large projection screens that are arranged in a symmetric, wide-angle configuration. Each projection screen measures 2.75 metres in width and 2.15 metres in height, leading to an effective display width of 8.25 metres across all three screens. At distances of approximately one metre or less, the three screens can immerse observers in a visual display environment that encompasses their effective horizontal and vertical fields of view.

Each projection screen is illuminated by a separate ceiling-mounted Epson PowerLite 7500C TFT active matrix LCD projector that forward-projects onto its assigned display screen. Each projector has a maximum display resolution of 1024 by 768 pixels and a maximum display refresh rate of 60 Hz. The projectors are rated to have an image brightness of approximately 800 lumens and support full 24-bit colour reproduction. The dynamic range of the projectors is not exceptional, which is most noticeable in an observable lightness in the black levels of the projectors. For our experiment, only the centre display of the three projection displays was used to render visual stimuli. The other two projection displays remained inactive, although



Not shown: Fastrak stylus, 3D stereo glasses, head tracker

Figure 3.1: Schematic Overview of the Experimental Virtual Environment

their presence was useful in maintaining a level of perceptual decontextualization that is reminiscent of many large scale, immersive virtual environments.

Aside from the main display area, a special workspace was co-located in the lab in order to provide an observation space for the supervising experimenter and to organize all of the computing hardware that was used to drive the experimental software that collected data from each subject. A desktop PC workstation provided the hardware for the presentation of visual stimuli and the processing of user interactions during the experiment. The PC platform was configured with an Intel Pentium 3 processor, driven at a clock speed of 700 MHz. The workstation had 384 MB of RAM and an Nvidia Geforce 2 GTS graphics card with 32 MB of video mem-

ory. A Creative Labs Sound Blaster series sound card and a pair of generic desktop stereo speakers provided audio functionality for the experiment. Through the use of a monitor signal switcher, video signals from the video card were presented to the appropriate LCD projector in the Landscape Immersion Lab and simultaneously to a 19" CRT display in the experimenter's workspace. A network interface provided access to networked data storage facilities. During experiment implementation and execution, the PC operated under Red Hat Linux, version 7.0.

Subjects were seated at a measured distance of 2.5 metres away from the centre projection display. Seating was carefully arranged so that the middle of the centre projection screen fell directly in line with subjects' physical midlines. At this distance, the centre projection display was measured to subtend approximately 63 degrees of subjects' horizontal field of view. A tall, wooden table with a width of 120 centimetres, a length of 80 centimetres, and a height of 95 centimetres was situated immediately in front of subjects. By ensuring that subjects kept their arms and hands underneath the table at all times during the experiment, the wooden table provided a sufficient mechanism for eliminating undesirable visual feedback. Immediately behind subjects, another wooden table held the pair of generic desktop speakers that conveyed audio to subjects throughout the experiment.

Although the presented visual stimuli provided no stereoscopic depth cues, subjects were required to wear a pair of inactive ASUS stereo glasses throughout the experiment. These glasses were present during the experiment in order to simulate conditions that are representative of other immersive virtual environments and in order to aid in the removal of unwanted contextual perceptual cues. In order to provide facilities for subject interaction during the experiment, a Polhemus Fastrak magnetic tracking device was employed. The Polhemus Fastrak is a six degree-of-freedom tracker that simultaneously accommodates up to four sensors and input devices of various kinds that permit spatial interaction along three positional axes and three rotational degrees of movement. In this experiment, two input sensors

were used in conjunction with the Fastrak: a passive receiver, which was mounted onto a plastic helmet that subjects wore during the experiment in order to provide information about subject head movements, and an interactive stylus with a push-button, which was used to provide subjects with a suitable implement for acquiring visually presented targets through aiming and pointing movements.

With the two input sensors attached to it, the Fastrak was capable of capturing the three-dimensional spatial positions and orientations of each input device at a rate of 60 Hz. Because the magnetic transmitter that was employed by the Fastrak had a maximum effective range of approximately three metres, with noticeably degraded performance at distances greater than about one metre, the magnetic transmitter was situated on a stable wooden base to the immediate left of subjects. Thus, the magnetic transmitter was always located at a distance of no more than one metre away from the Fastrak input devices. All reasonable efforts were made to ensure that metallic objects and any items with a considerable metal component were kept away from the vicinity of the magnetic transmitter and input devices, thereby preventing most metallic interference that could have disrupted the functionality of the Fastrak.

A custom software application integrated all of the experimental apparatus into a fully-functional, cohesive system. The software provided a dynamic organizational engine for controlling the progress and flow of the experiment. It also provided a facility for presenting visual stimuli and permitting subject device interaction with very specific timing parameters, drivers and configuration tools for employing and calibrating the projection display and Polhemus Fastrak, and a persistent mechanism for recording subject responses during each experiment session. For implementation, the Java programming language and software development kit (SDK), version 1.4 were used. Support for rendering graphics was provided by the Java Abstract Windowing Toolkit (AWT) and the Java 2D API.

3.3 Experimental Presentation

We defined our experiment in terms of a unique set of trials that were consistently repeated in each experimental condition so that we could assess the similarities and differences between the performance characteristics of each interaction style. Such a design permitted us to determine whether or not there was a behavioural difference between verbal reporting and motor reporting of targets that exhibited an induced Roelofs Effect, as well as any other differences between the four experimental conditions.

All of the previous two visual systems experiments that used an induced Roelofs Effect specified a finite number of discrete target and frame positions as parameters that were used to make up a repeatable set of trials. In our experiment, we followed a similar strategy by specifying a series of experiment trials that could be carried across all of the experimental conditions. Trials were specified in terms of four different parameters: target position, frame position, response delay, and trial repetition.

The first two parameters of target position and frame position went hand in hand because the properties of the induced Roelofs Effect needed to be considered when adapting the visual illusion for use in this experiment. In particular, there were issues about the visual shape and size of the frame and target that were addressed. Consistent with past studies of the two visual systems hypothesis, we selected a frame and target with dimensions that took up a considerable portion, although not the entirety, of a subject's field of view. Through an examination of previous two visual systems experiments, we chose to use circular targets and rectangular frames with varying positions in order to create our version of the induced Roelofs Effect. These objects were flat shaded in the solid colour tones of red and green respectively.

Our circular targets were 0.5 degrees of visual angle in diameter. These circular targets occupied one of three discrete positions in the presented visual space of the subjects; the centremost target fell directly on the physical midlines of sub-

jects, while the centrepoint of each of the other targets were spaced at distances of 1.5 degrees apart from the centre. The expected effect of the induced Roelofs Effect suggested that subjects would misperceive two additional target positions in their visual space in addition to the three defined positions, although these were not explicitly present in any of the actual experiment trials. Through the induced Roelofs Effect, targets may have appeared to be present in additional positions up to three degrees away from the centre target. It could be argued that these two additional illusory positions should have been added to the list of explicitly presented target positions in order to ensure that subject perceptions of these target positions in non-illusory trials were consistent with the perceptions of other target positions. However, previous studies of the two visual systems hypothesis have noted that the presence of these two additional positions in presented trials is unnecessary because nearly all variance in subject responses as a function of target position can be reconciled by a linear regression (Bridgeman, 2000). Thus, we safely eliminated these two additional target positions from the set of presented trials, thereby reducing the total number of trials that needed to be presented to subjects.

We made our rectangular frames to be 21.0 degrees in width and 9.0 degrees in height. These frames also had a line thickness of 1.0 degrees. In order to create an induced Roelofs Effect, these rectangular frames occupied one of three discrete positions in the presented visual space of the subjects. A non-offset position, where the centre of the rectangular frame coincided with the midlines of subjects, created a control setting for the induced Roelofs Effect. When the rectangular frame was in this position, subjects were expected to be able to accurately perceive the locations of targets relative to themselves. Two other positions, where the centre of the rectangular frame was offset four degrees to the left or right of the midlines of subjects, was expected to create the misperception of target locations that is representative of the induced Roelofs Effect. Figure 3.2 provides a simultaneous presentation of all of the target positions, including the two expected misperceived target locations,

and the non-offset frame position.

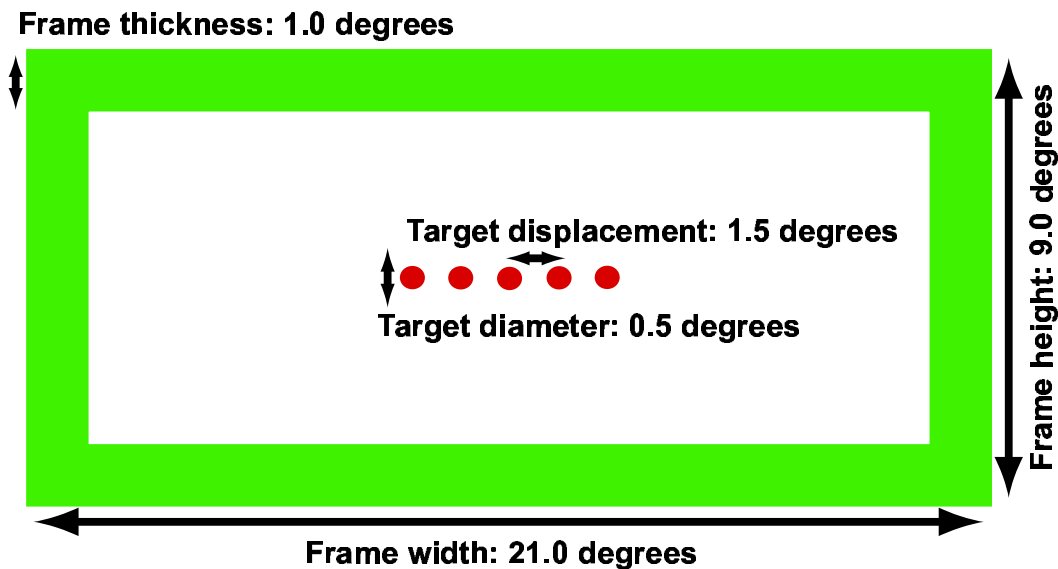


Figure 3.2: Frame and Target Dimensions for the Presented Visual Stimuli

The two other parameters were response delay and frame repetition. The parameter of response delay was used to observe the time-limited nature of the sensorimotor stream in virtual environments. For this reason, this parameter determined the amount of time that elapsed between presentation of a target and frame and permission for subjects to begin their responses. There were two different response delays: a zero second delay and a four second delay. With the zero second delay, it was expected that subjects would provide accurate motor responses, while at the four second delay, it was expected that subjects would begin to provide more inaccurate motor responses in the open loop pointing condition. The parameter of trial repetition determined the number of times that a trial with a particular target position, frame position, and response delay was repeated. Such repetitions were necessary in order to ensure the consistency of subject responses. However, an important factor in selecting an appropriate number of repetitions was the determination of a reasonable total number of trials. Since we wanted to avoid experimental obstructions such as subject boredom and fatigue, we specified a total

of three repetitions for each trial type. Table 3.1 shows the set of trials across all four parameters. There were 54 distinct trials in each condition of the experiment, which led to a total of 216 presented experimental trials.

	Target Positions	Frame Positions	Response Delays	Trial Repetitions
Number of	3	3	2	3
Total Experiment Trial Types: 54				
Across All Four Conditions: 216				

Table 3.1: Total Number of Trial Presentations in the Experiment

During the experiment, trials were blocked by condition so that a subject would be presented with the same set of trials under each of the different interactive conditions. Under each of these blocked conditions, each set of 54 trials was randomized in order to avoid any ordering effects. Each of these trials was presented to subjects in roughly the same manner across all of the conditions. For any given trial, subjects were simultaneously presented with a circular target in one of the three target positions, and a rectangular frame in one of the three frame positions for a period of exactly one second. After this one second period elapsed, the target and frame pair were extinguished from the projection screen, leaving subjects with a black display. Subjects were subsequently given either a zero or four second response delay, after which subjects were prompted with an audio request to provide a verbal or motor response, depending on the condition that they were currently completing. Subjects provided their verbal or motor response, and in the motor response conditions, were provided with an opportunity to move their limbs into a resting position prior to starting the next trial. No equivalent opportunity was provided for subjects in the cognitive report condition because such movements were unnecessary for verbal responses. There was a two second delay that separated trial presentations from one another. During this delay, the experiment software application wrote all of the observational data for the trial into a text file for later analysis. The experiment application recorded information about the properties of the presented trial, includ-

ing the trial number, target position, frame position, response delay, and level of interactive latency. It also recorded information about the response that subjects provided, the head position and orientation of subjects at the time of response in the trial, and the amount of time that it took for subjects to provide a response, starting from the time that they were prompted to provide a response to the time at which the software application detected a response.

3.3.1 Subject Consent and Preparation

Sessions with experiment subjects during the execution of this experiment involved a progression through three separate phases of tasks. The first phase necessitated subjects to provide written consent and to become set up to complete the tasks specified by the experiment. The second phase required subjects to complete the four outlined experimental conditions. The third phase provided subjects with monetary compensation and a verbal debriefing that disclosed the motivation and reasons for requiring their participation. An experimenter was always present throughout the course of an individual experiment session.

In order to satisfy the ethical requirements surrounding behavioural experiments set by the Office of Research Services at the University of British Columbia, participating subjects were asked to provide written acknowledgement that they agreed to participate in this experiment. For this purpose, consent forms with a general description of the experiment were provided. If subjects agreed to the experiment, they provided their written signature on two forms. One form was kept by the experimenter as a permanent record of the experiment session, and another form was given to the subjects for their own records. Once this was done, subjects were directed by the experimenter to the experiment presentation area, where they were prepared for the next phase of the experiment by asking them to wear the prescribed pair of stereo glasses and plastic helmet with the mounted Fastrak receiver. Because the first condition involved verbal responses, the Fastrak stylus device was

not given to subjects at this point. After the set up was complete, subjects were asked if they had any questions before the experiment began. If there were any issues, the experimenter provided answers to any questions that they might have had. If there were no questions, the experimenter informed them that they were about to receive a set of audio instructions that would provide them with further information about what was required from them.

The experimenter indicated the start of the second phase of the experiment by extinguishing all of the ambient light sources in the lab. Thus, with the exception of illumination from the LCD projector and projection screen, the lab was made completely dark. The experiment software application was started, and a set of streaming audio instructions were provided via the desktop stereo speakers to subjects. These instructions provided more general information about the experiment, specifying more directly what kinds of actions were expected of them throughout the session. These audio instructions included an estimate of the duration of the session, directives to listen to all of the audio instructions carefully, and a reminder of a subject's freedom to withdraw from the experiment at any time. After these audio instructions were played, subjects were required to give a verbal confirmation that they were ready to proceed with the experiment. Upon this confirmation, the experimenter moved the experiment software application into the first of the four experimental conditions.

3.3.2 Cognitive Report Condition with Verbal Responses

The first of the four experimental conditions to be presented to subjects was the cognitive report condition. Subjects received further audio instructions that indicated the nature of the visual stimuli that they were about to receive and the actions that were required of them. Specifically, subjects were told that they were about to undergo a series of experimental trials that would involve the presentation of circular targets and rectangular frames, and that for each trial, they would be re-

quired to make a verbal response. These verbal responses would indicate where the subjects believed targets were, in relation to their perception of straight ahead. Subjects were also directed to keep their heads as steady as possible and to keep their arms underneath the wooden table frame at all times. Once these audio directions were presented, subjects were provided with an opportunity to ask any questions regarding what was expected of them.

If they had no questions, then they were presented with a labeled description of the five supposed positions at which targets were presented. In reality, there were only three positions that were presented in the actual experimental trials, but subjects were led to believe that there were five positions so that they could provide appropriate responses in the case that they misperceived target positions because of the induced Roelofs Effect. Moreover, in the subsequent practice trials for this condition, subjects were presented with each of the five displayed target positions. This was done in order to ensure that subjects would not form a strategy of selecting only three of the five possible target choices during the experimental trials. Text labels and audio notes indicated the responses that they should give if they believed that targets were in a specific location. Thus, if a target appeared to be presented in alignment with their midline, they should have responded with the answer “Centre.” Other responses followed from where the target appeared to be, relative to their midline. Subjects may also have provided responses of “Far left,” “Left,” “Right,” or “Far right.” In order to provide responses to the experiment software, the software was designed to assign the mnemonic keys 1 to 5 with values corresponding to the labels. Thus, when subjects provided verbal responses, the experimenter could press appropriate keys in order to notify the experiment software of the provided responses.

When subjects indicated that they were comfortable with their knowledge of the labels and target positions, the software provided subjects with a series of practice trials that permitted them to become accustomed to the verbal task that they

were expected to perform. In order to provide subjects with an experience that was as seamless as possible, subjects were not explicitly told that they would be provided with practice trials. Instead, they were first told that they would undergo a series of trials that only involved the presentation of targets with verbal feedback. These trials proceeded in the same manner as typical experimental trials, except that no frames were presented and subjects were provided with audio feedback that indicated the correctness of their responses. These target-only presentations continued until subjects completed a minimum of ten trials, and they demonstrated that they were able to correctly identify the target positions of five targets in succession.

Subsequently, subjects were told that they would receive a series of trials that involved the presentation of both targets and frames, while continuing to receive verbal feedback. In these trials, only the target position changed from trial to trial, while the frame remained locked in the non-offset position. Just as in the previous series of trials, subjects advanced only after they completed a minimum of ten trials, and they were able to correctly identify the positions of five targets in a row. At this point, subjects were notified that they would continue to receive these target and frame presentations, with the modification that they would no longer receive any feedback regarding the correctness of their responses. Subjects were provided with at least ten more trials, advancing without any additional notice to the experimental trials when they again demonstrated an ability to correctly identify the positions of five targets in a row. As before, the frame was always in the non-offset position; only in the experimental trials did the frame also assume one of the two non-offset positions.

Progress through this last set of practice trials and the actual experimental trials proceeded in the manner previously described for typical trial presentations. Completion of the actual experimental trials was denoted by an audio indication to the subject that the cognitive report condition was complete, and that they could take an opportunity to relax. At this point, subjects were provided with an

opportunity to make comments or ask questions. During this rest period, the lab continued to remain dark and subjects remained seated, continuing to wear both the stereo glasses and head-tracking helmet. There was no predefined bound for the duration of this period, and subjects only advanced to the next condition when they indicated that they were ready to do so.

3.3.3 Motor Report Conditions with Pointing Responses

When subjects indicated their willingness to move forward to the next condition, the experimenter provided them with the Fastrak stylus in order to complete the rest of the experimental conditions. The presentation of trials in the remaining motor response conditions varied only slightly from their presentation in the cognitive report condition. In order to ensure that subjects did not exhibit any effects of interaction between the open loop pointing condition and the closed loop pointing condition, these two conditions were counterbalanced such that half of the participating subjects experienced the open loop pointing condition first, followed by the closed loop condition. Thus, the other half of the participating subjects were presented with the closed loop pointing condition first, followed by the open loop condition. Counterbalancing conditions in this way provided an additional assurance that subjects were exhibiting a true dissociation of the cognitive and sensorimotor streams, and that any observed effects were not the result of developing a memory-based strategy in one condition that altered the completion of a successive condition. All of the subjects experienced the closed loop pointing with interactive latency condition as their last experimental condition.

The progression of the actual experimental trials proceeded in exactly the same way that trials were presented in the cognitive report condition, with the exception that subjects were expected to respond with a motor pointing response instead of a verbal response. Moreover, subjects were provided with no visual feedback of any kind while they were making responses in the open loop pointing condition.

However, during the closed loop pointing and closed loop pointing with interactive latency conditions, subjects were provided with a white aiming crosshair that was one degree wide by one degree high and that was designed to provide correspondence between subject aiming motions and projected virtual pointer positions on the screen.

In all cases, subjects indicated their motor responses in exactly the same way. Upon receiving an indication to respond, subjects used the provided stylus to point at the position where they believed a circular target had been presented. When subjects were satisfied with their aiming response, they held their pointing position and orientation until the experiment software provided an audio indication that the trial was complete. This audio indication came precisely two seconds after the subjects had settled into a steady pointing position. At this point, subjects returned the stylus to a resting position on their laps and pressed the button on the stylus when they were ready to proceed with the next trial in the current experimental condition. At first glance, such a “point-and-dwell” response mechanism appears unintuitive, given that we might expect the stylus to work in the same way that a mouse does. However, as a result of unintended pointer movement during a button press, such a “point-and-click” response mechanism is susceptible to acquisition errors (Olsen & Nielsen, 2001). Thus, by using a dwelling response, we were likely to receive better data regarding the motor responses of subjects.

The most important differences among each of the remaining experimental conditions came from the presentation of practice trials prior to the actual experimental trials. Similar to the cognitive report condition, each of the motor conditions provided a seamless progression of trials from a set of target-only trials with verbal feedback, to a set of target and non-offset frame trials with verbal feedback, to a set of target and non-offset frame trials with no verbal feedback. In all of the motor response conditions, the set of target-only trials with verbal feedback and the set of target and non-offset frame trials that followed were also initially accompanied by

visual feedback in the form of the aiming crosshair that was used in both closed loop pointing conditions. In the open loop pointing condition, these first two sets of trials were followed by a set of target and non-offset frame trials that removed the aiming crosshair but retained the verbal feedback, and then by a set of target and non-offset frame trials that presented no visual or verbal feedback of any kind. In both of the closed loop conditions, these first two sets of trials were immediately followed by a set of target and non-offset frame trials that removed the verbal feedback, but retained the aiming crosshair. Furthermore, in all of the motor response conditions, exactly five pointing trials were provided for each progressive set, meaning that there was a total of twenty practice trials for the open loop pointing condition and fifteen practice trials for each of the two closed loop pointing conditions.

In the closed loop pointing with interactive latency condition, the aiming crosshair did not lag until the actual experiment trials. When the aiming crosshair exhibited lag, it did so by thirty frames. This meant that the position of the aiming pointer was consistently half of a second behind where the position of the aiming pointer should actually have been. Thus, there was significant discord between the aiming movements provided by subjects and the projection of their aiming movements in the virtual environment. Although this amount of lag is improbable in even the most latency-plagued VR systems, such an overpowering sense of lag ensured that there was no noticeable correspondence between subjects' aiming motions and what information was provided by the aiming crosshair. In effect, this permitted us to observe the influence of lag on pointing performance with a much smaller group of subjects than might otherwise be necessary. By employing a larger amount of lag, we expected to exaggerate the size of the perceptual effects that we were interested in observing and to have made more subjects susceptible to these effects so that we were able to assess the influence of lag without having to use a much larger number of subjects.

3.3.4 Experiment Session Debriefing and Subject Compensation

Following the completion of all of the experiment conditions, audio from the experiment software indicated to subjects that the experiment was complete. The experimenter helped subjects in the removal of the experimental apparatus and normal lab lighting was restored. For their participation in the experiment, subjects were provided with a twenty-five dollar honorarium. Subjects were required to complete two copies of a receipt form in order to acknowledge their receipt of this honorarium. One copy was retained by the experimenter as a permanent record of the transaction, while the second copy was provided to the subject for their own personal records.

Subjects were also provided with an experimental debriefing that outlined the general purpose and motivation for this experiment. The debriefing included a short overview of the two visual systems hypothesis, and the hypotheses that the experiment set out to test. If subjects had any final questions, they were free to ask the experimenter at this point. Furthermore, the experimenter encouraged subjects to provide constructive comments about the experiment session that they completed. These comments were noted for later reference in the subsequent analysis and presentation of subject data.

3.4 Implementing the Experiment

The process of developing the software application that coordinated the experiment involved a significant number of implementation details that may prove useful in future research, regardless of its relevance to this current work. In particular, there are at least four different areas that may be of interest. First, there are the details surrounding the implementation of audio instructions in the experiment. Second, there is the derivation of appropriate techniques for rendering objects with a desired visual size. Third, the details of deploying the Polhemus Fastrak as a virtual pointing

device provide information that does not appear to be widely reported elsewhere. Fourth, information on calibrating and integrating control for inducing interactive latency in the Polhemus Fastrak may be useful as a starting point for those who are interested in performing similar work with spatial tracking devices.

3.4.1 Using Audio to Convey Instructions to Subjects

Pre-recorded audio instructions and audio notation, as implemented in this experiment, do not appear to be commonly implemented elsewhere. Although experimental studies that involve normal human subjects are common in experimental psychology, HCI, or many other fields involving human behaviour, the use of pre-recorded audio in experiments is not widely reported. In most interactive studies, instructions are more likely to be conveyed to subjects through pre-written text, or through the voice of a supervising experimenter as instructions and notes that are read off of a prepared script. Although the earliest designs of this experiment included the use of a traditional text-based approach to instructions, the final design of the experiment incorporates pre-recorded audio instructions for at least three main reasons.

First, the context of virtual reality provides an excellent environment that facilitates the presentation of experiment instructions and notes in an auditory form. For this particular experiment, the design and editing of a suitable recording script is relatively straightforward, and the recording process is simpler in many ways when compared with the many decisions, such as text format and readability, that would have to be made in rendering text instructions for subjects to read. The implementation of audio in the experiment software is also relatively straightforward because a considerable amount of support exists in the Java API for reading and playing back pre-recorded audio files. Second, pre-recorded instructions remove a considerable amount of effort from the process of running an experiment session. Because experimenters are no longer required to recite the same set of instructions to the

many subjects that they will be supervising, experimenter fatigue is substantially reduced and experimenters can focus more of their energy and attention on observing subjects. Third, our subjects appeared to be quite receptive to pre-recorded audio instructions, and they seemed to understand their instructed tasks well. In the current experiment, subjects seemed to appreciate the presence of pre-recorded audio instructions because there was less of a chance of misunderstanding what was expected of them during the experiment.

The audio instructions in this experiment were recorded in a WAV/PCM format at 16-bits per sample and 22500 Hz. Although any of a number of different audio formats and encoding qualities might have been sufficient, this particular format provided an excellent balance between ease of implementation, audio quality, and storage requirements. With audio at this level of quality, any noticeable background noise as a result of recording and encoding was eliminated, and individual recordings could be relatively lengthy without using up an enormous amount of disk storage space. This particular experience with recording audio for use as instructions for a software application has also led us to the development of a few guidelines that might be helpful in future implementations.

Preparation of a completely edited text script and organization of audio files prior to recording is critical to successfully using audio. By knowing where and when audio is going to be used in the software application, and by knowing what is going to be recorded, the implementation of an audio instruction component in a software application is likely to be much faster. A proper text script that has been thoroughly proofread helps to ensure that the audio that is being recorded is less likely to cause problems with the subjects that are required to listen to it. Additionally, it is important that recorded words are spoken slowly and clearly. In doing so, the resulting audio files are less likely to be garbled, and subjects will have more of a chance to understand the instructions that are being given. Finally, using a good quality microphone and ensuring that the recording environment is noise free

is also vital. The quality of the microphone that is used has a large influence on the overall quality of the audio that is produced. Keeping the recording environment silent, with the exception of the instructions that are actually being recorded, helps to ensure that the audio that is intended to be heard is not overpowered by other audible artifacts.

3.4.2 Determining the Visual Angle of Visible Objects

Because measuring visual stimuli according to an absolute metric, such as centimetres or inches, tells us nothing about the proportion of the visual field that is taken up by presented stimuli, the presentation of visual objects is usually measured in terms of relative *visual angles* in psychological experiments. By referring to visual stimuli in terms of their angular size, it becomes possible to more accurately replicate experimental conditions in future research and it also provides some information about the observational conditions of the current experiment. Furthermore, if we are to render objects in a virtual environment that are of some consistent size to all subjects in an experiment, a knowledge of visual angles is critical. Fortunately, the mathematics behind determining the visual angles of visible objects is not particularly difficult, only requiring that we have a basic knowledge of trigonometry and algebra.

Figure 3.3 describes the basis for the trigonometry that leads to formulas for calculating the visual angle of objects. Observe that the visual angle of a given visual stimulus is a function of the distance of a viewer from the visual stimulus and the actual physical size of the visual stimulus. In addition, observe that this given distance also conveniently divides the subtended visual field of a viewer into two right triangles of equal size. We assume that the objects of interest are 2D projections onto a planar display and that a symmetric, linear approximation of the visual angles to be used is sufficient for the task at hand. More formally, we let θ be the visual angle of a given visual stimulus in degrees, we let D be the distance

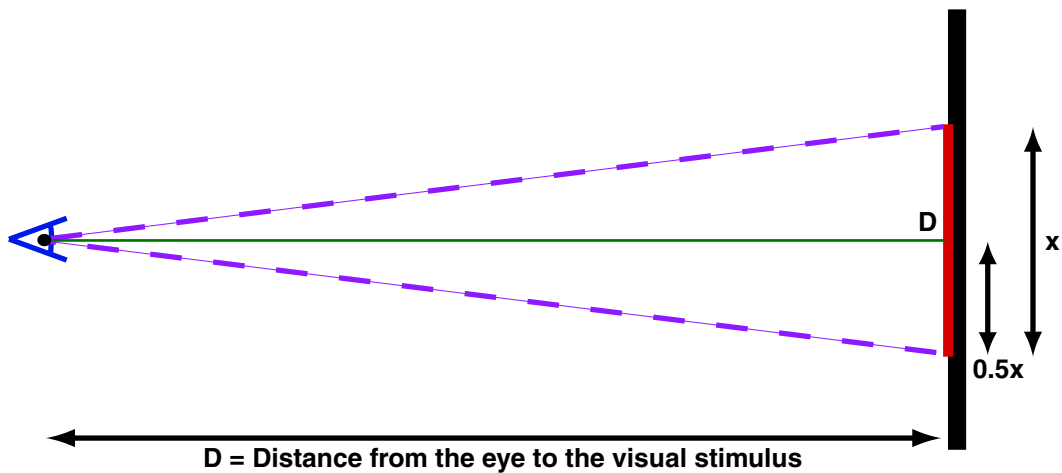


Figure 3.3: Determining Visual Angles with Trigonometric Techniques

between the eye and the visual stimulus, and we let x be the size of the visual stimulus in some consistent unit of measurement. Then:

$$\tan\left(\frac{\theta}{2}\right) = \frac{x}{2D}$$

Thus, in order to find the visual angle of a stimulus, we need to solve for θ , which can be done given that we have values for x and D :

$$\theta = 2\left(\arctan\left(\frac{x}{2D}\right)\right)$$

Rewriting this equation in terms of θ and viewing distance D yields a formula that can be used in order to determine the correct physical size of an object of a desired viewing size:

$$x = 2\left(D \tan\left(\frac{\theta}{2}\right)\right)$$

Note that in order to obtain enough information to render the object, we will need to do this calculation twice: once for the visual width of the object, and once for the visual height of the object. Once the physical measurements of the desired visual object have been determined, rendering the object on a display screen requires the conversion of the physical measurement into pixel units. Such conversions require

that we know the physical dimensions of the display and the pixel resolution at which the display is being rendered. Let U be the number of pixels per single unit of measurement. Let L be the size of one side of the display in some consistent unit of measurement, and let P be the corresponding pixel resolution of the display side measured by L . Then:

$$U = \frac{P}{L}$$

Moreover, if we let R denote the number of pixels that are required in order to render one dimension of an object at a given visual angle θ , then:

$$R = Ux$$

Thus, performing this final calculation for both dimensions of the desired object provides us with the information that we need for graphical rendering on a planar display of fixed size.

3.4.3 Using the Polhemus Fastrak as a Virtual Pointing Device

For at least three reasons, the Polhemus Fastrak is a convenient tool for implementing a simple virtual pointing device based on ray casting techniques. First, a lightweight stylus device is available as a sensor option for the Fastrak, providing a tool that has the weight and feel of an archetypical pointing implement. Second, the Fastrak provides tracking for both spatial position and orientation, meaning that it is possible to simulate the behaviour of other pointing solutions, such as laser pointers. Third, good documentation, a simple hardware interface, and programming support make implementation with the Fastrak a reasonably straightforward process.

In order to implement a virtual pointing device with the Fastrak, an external API that links the Fastrak device with a host machine must be implemented. For this current experiment, a multithreaded programming interface with Java was established through a serial port connection between the Fastrak and the experiment

PC workstation. In this configuration, the Fastrak continuously transmits records of the current spatial position and orientation of the stylus through the serial port connection, which are subsequently processed in a background thread that takes these data and uses them in order to specify a projected position on the screen. This projected position may be rendered as a crosshair, mouse pointer, or whatever pointer representation is desired.

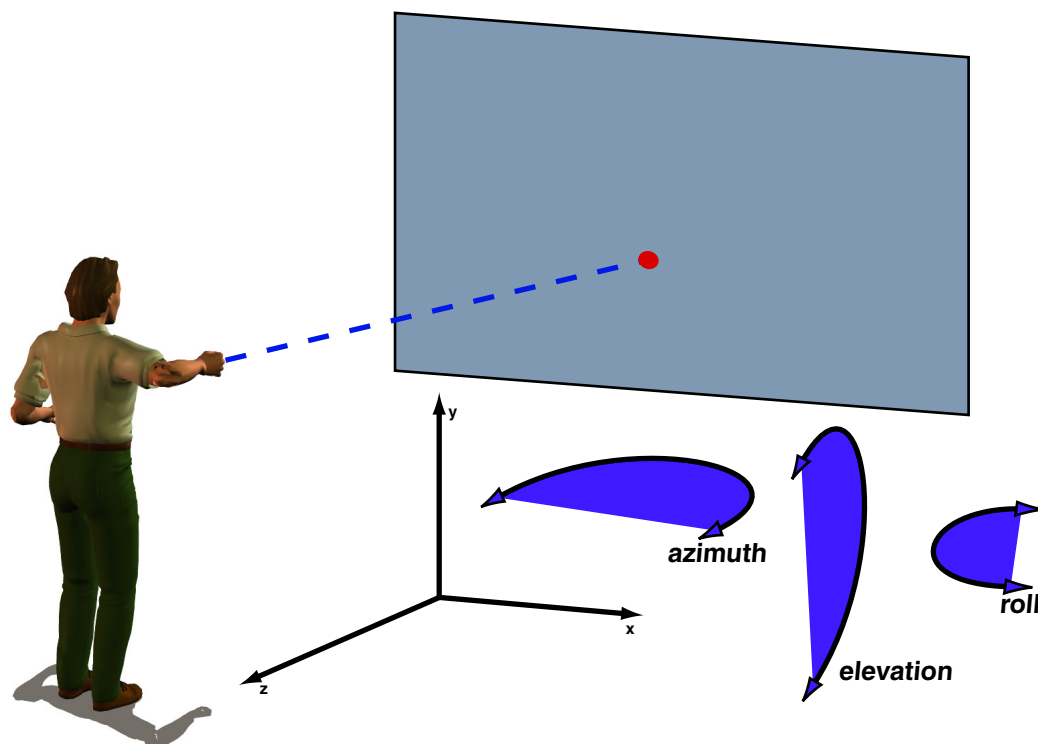


Figure 3.4: Virtual Pointing Configuration with Ray Casting

An issue of significant concern is developing an appropriate correspondence between three dimensional spatial coordinates and two dimensional screen position. A simple ray casting algorithm is used in order to perform the appropriate projection. Figure 3.4 presents an example of virtual pointing with ray casting, identifying the three positional axes and the three orientational axes that are measured by the Fastrak. For each spatial record that is retrieved from the Fastrak, we extract the

provided position and orientation information. Let us define a sextuple T such that:

$$T = \{x, y, z, azimuth, elevation, roll\}$$

Thus, T defines the position and orientation information stored in a single Fastrak record. In order to perform ray casting, we define X_{proj} and Y_{proj} to be the coordinates that specify the screen projection point given by intersecting the virtual “ray” that originates from the stylus device with the planar display screen. Using some straightforward trigonometry, these coordinate values can be determined by observing that the coordinates are simply a function of the distance and orientation of the pointing device. Therefore:

$$X_{proj} = \tan(T_{azimuth})|T_z|$$

$$Y_{proj} = \tan(T_{elevation})|T_z|$$

With these equations, the resulting coordinate points can be calculated and converted into the appropriate pixel locations and then rendered on the display screen as required. In the present experiment software, these coordinate points are translated into mouse coordinates, demonstrating that the Fastrak and attached stylus device can be effectively used as an alternative pointing device for mouse-style pointing interactions. The “point-and-dwell” mechanism that is used in the experiment can be implemented by maintaining a list of recently retrieved spatial records and calculating the standard deviation for each of the positional and orientational values in the list at some regular time interval. When all of these values are below a definable threshold value, a dwelling interaction is said to have been accomplished. Based on pilot observations of motor responses and other studies of pointing interactions, it appears that determining an appropriate standard deviation threshold is likely to be dependent on subject hand jitter and distance from the display of interest (Myers et al., 2002). We used this technique of calculating standard deviations to establish our own “point-and-dwell” mechanism.

3.4.4 Calibrating and Lagging the Fastrak

In the present experiment, use of the Polhemus Fastrak required that some attention be devoted to the processes of calibration and registration in order to ensure that the data that are acquired from the tracker are accurate and reliable. It was deliberate that the physical configuration of the virtual environment permitted the placement of the Fastrak magnetic transmitter within very close proximity to the sensor devices that were used in the experiment and that the environment was free of any significant sources of magnetic interference. As such, many of the complex filters and compensatory measures that might otherwise be necessary in a larger space were avoided in the current implementation without too much of an impact.

Measurements of the Fastrak sensors in a stationary position within the virtual environment suggested that the sensor jitter due to distance and magnetic interference created error within two screen pixels at the subjects' distance of 2.5 metres. Alignment of the Fastrak stylus device and the onscreen crosshair employed in the experiment was accomplished through careful visual inspection and adjustment. The result of the alignment procedure was a correspondence between crosshair and stylus position with no distinguishable offset. Although significantly more complex alignment techniques could have been used, the performance provided by the visual inspection method was deemed sufficient for this particular experiment and implementation.

The simulation of interactive latency was accomplished through the implementation of a circular queue data structure of spatial records from the Fastrak. Data records from the Fastrak were added onto the end of the queue, and the oldest available records expired at the front of the queue. In order to induce controllable lag in the virtual crosshair that was being presented, we simply maintained a pointer that acts as an index to the supposed "most recently seen" Fastrak record. Because old records were constantly being removed from the circular queue and were being replaced by newer records, the position indicated by the pointer was also constantly

being updated. The newest records in the queue were not used to update the virtual crosshair until they moved forward, into the queue position indicated by the pointer. Thus, when the pointer was set to point to the end of the queue, we updated the virtual crosshair with no induced latency because we were always “seeing” the most recently received record. We increased the amount of lag that we introduced into the system when the pointer moved progressively closer toward the front of the queue because it took an increasing amount of time for the most recently received record to be “seen” by the experiment application.

Chapter 4

Experimental Results

A number of interesting results and effects were observed in the statistical analysis that was conducted subsequent to the design and implementation of the two visual systems experiment in VR. All of the collected data from subjects that participated in the experiment were processed offline using a variety of statistical measures and techniques. Descriptive statistics were compiled in order to summarize the quantitative measures that were taken between subjects. Two-way analyses of variance (ANOVAs) over the independent variables of target position and frame position were conducted. These two-way ANOVAs were run for each subject, each experimental condition, and each response delay. Supplemental statistical methods were used in order to further characterize the performance of subjects in ways that extend beyond the results of the two-way ANOVAs across all of the experimental conditions.

4.1 Summarizing the Results

Because we are primarily interested in looking for evidence of a dissociation between the two streams of visual perception, we can determine whether the two visual systems hypothesis applies to virtual environments by examining the extent to which there is a presence or absence of the visual illusion that was used in order to induce a misperception of object positions. In this present experiment, this reduces to

the presence or absence of the induced Roelofs Effect for each subject in each of the experimental conditions. The general criterion for determining the existence of induced Roelofs Effects for a subject was chosen to be the existence of a main effect of frame position at a significance level of less than or equal to 0.05. This tells us whether subjects' responses of target position varied significantly due to placement of the frame in each of the different response modes and response delays. For the cognitive report and the closed loop pointing conditions, we expect that subjects will exhibit an induced Roelofs Effect when permitted to respond immediately, while for the open loop pointing and closed loop pointing with interactive latency conditions, subjects do not exhibit the effects of such a visual illusion when permitted to respond immediately. A summary of the most prominent results arising from our statistical analyses is presented below. The subsequent subsections in this chapter describe specific results and the individual analyses in more detail. In all of the analyses that follow, we adopt the convention of reporting the smallest F ratios for significant results, and the largest F ratios for non-significant results.

Across all four conditions, all fourteen subjects consistently demonstrated highly significant main effects of target position [$F(2, 18) > 6.78, p < 0.006$]. This indicates that subject responses differed significantly in relation to changes in presented target position. Thus, when targets were presented toward the left, subjects had a significant tendency to provide responses that were more to the left, and when targets were presented to the right, subjects had a significant tendency to provide responses that were more to the right. Observing the reliability of these main effects is especially important to our overall experimental analysis. They confirm that subjects consistently provided responses in relation to what was presented in the experiment, and that subjects were not likely to make random responses to presented targets. If these particular main effects were not present, we would not be able to rely on the results that followed from our other subsequent analyses of subject data.

A convenient summary of subject performance and the presence of the in-

Subject Number	Cognitive Report Condition	Open Loop Pointing Condition	Closed Loop Pointing Condition	Closed Loop Pointing w/ Latency Condition
1				
2				
3				
4				
5	● ○		●	
6	● ○	●		●
7	● ○		● ○	
8	● ○		● ○	● ○
9	● ○	●	● ○	
10	● ○		● ○	● ○
11	● ○	●	● ○	●
12	● ○		● ○	● ○
13	● ○		● ○	
14	● ○	●		●

- Observed induced Roelofs Effect in terms of a main effect of frame position
- Observed induced Roelofs Effect in terms of absolute number of trials showing effect

Table 4.1: The Induced Roelofs Effect Across Subjects and Conditions

duced Roelofs Effect is provided in Table 4.1. For each subject and each experimental condition, the table indicates which subjects exhibited the effects of the visual illusion and where these effects were demonstrated, in terms of whether effects were indicated by the initial two-way ANOVAs, by the alternate criteria of trial frequency, or by both analyses. One of the features of the table is its illustration of the relative sensitivity of each analysis method; although the results of both methods are reliable and consistent throughout all of the conditions, assessment of perceptual influence appears to be more prominent in the two-way ANOVAs than the alternate frequency-based criteria. Another feature of the table is its description of individual subjects' frequency-based performance across all of the experimental conditions. In this sense, the table permits us to see that subjects were most likely to fall under the influence of the induced Roelofs Effect in the cognitive report and closed loop pointing conditions, while they were far less likely to exhibit such effects in the open loop pointing and closed loop pointing with interactive latency conditions.

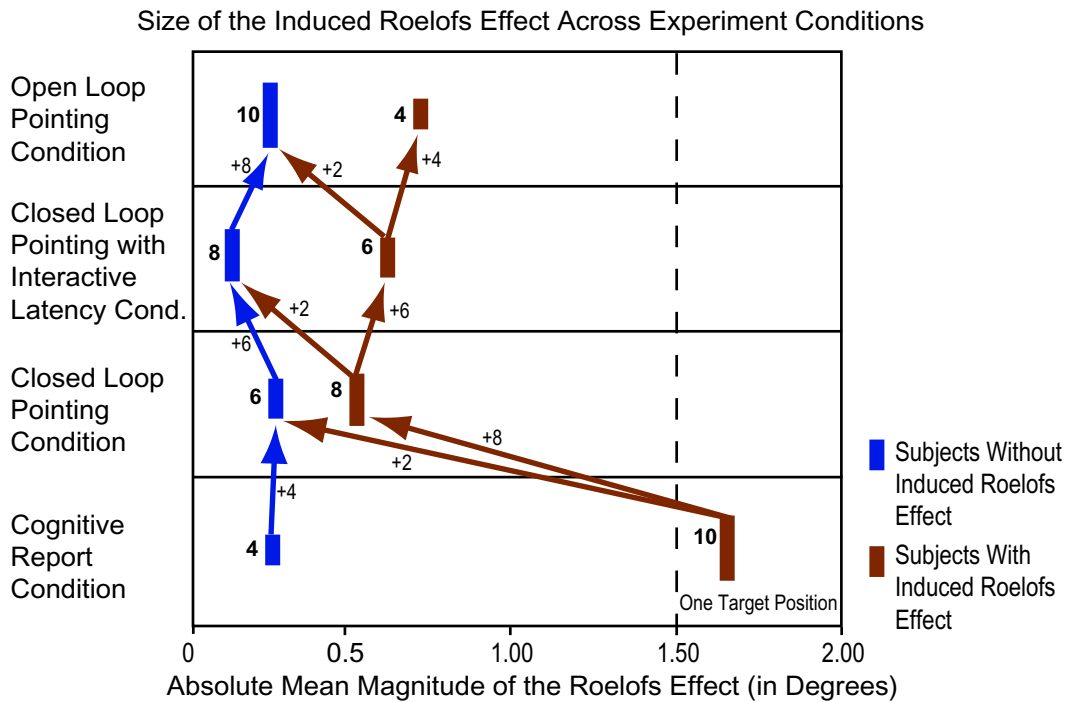


Figure 4.1: Influence of the Induced Roelofs Effect Across Conditions

Figure 4.1 also provides an interesting description of subject performance across all of the experimental conditions that is based on the measured magnitude of the induced Roelofs Effect. The figure provides an illustration of the number of subjects that either do or do not exhibit the induced Roelofs Effect in each of the conditions, showing where subjects make crossovers from one group to another. The figure clearly illustrates that there is a distinct division between those subjects that demonstrate an effect and those subjects that do not demonstrate an effect. Subjects in the former group were shown to have readjusted mean magnitudes of illusory effect size that were seen to stray farther from the actual target position in a direction consistent with the offset frame and illusory effect than those subjects in the latter group. Moreover, the figure shows that there is a movement of subjects from one group to another across conditions. This change in size between each group in each condition is indicated by the height of the vertical bars in each

condition. The connecting lines that are labeled with positive integers indicate the number of subjects that moved from the group that exhibited the induced Roelofs Effect to the group that exhibited no induced Roelofs Effect across each condition. Interestingly, we can observe that the movement of subjects only occurs in one direction: subjects do not move back and forth between groups in each of the conditions. Rather, they become decreasingly dependent on their cognitive representations of space across decreasing levels of visual feedback, and rely more on their sensorimotor representations of space.

4.2 Cognitive Report Condition

During the cognitive report condition, ten of the fourteen subjects demonstrated a main effect of frame position [$F(2, 18) > 4.46, p < 0.027$], suggesting that these subjects experienced the perceptual effects of an induced Roelofs Effect. The mean magnitude of the induced Roelofs Effect was 1.67 degrees, and the standard error was measured to be 0.26 degrees. For the four subjects that showed no main effect of frame position, the mean magnitude of their responses was 0.25 degrees, and the standard error was measured to be 0.41 degrees. Only three of the subjects who showed a main effect of frame position were male, while all seven female subjects demonstrated the effect. All of the observed main effects of target position and frame position were consistent across each of the two response delays.

Figure 4.2 presents profile plots that represent the estimated marginal means for two of the subjects in the cognitive report condition. The solid lines indicate the marginal means across each of the presented target and frame positions. Had a subject demonstrated a capacity to correctly specify target positions across all of the different frame positions, the marginal means would have shown no indication of slope. Thus, the dashed lines in the plots indicate the “optimal” marginal means for subjects. The subject responses in the plot are measured in degrees relative to the centre target position. As such, 0 degrees represents the centre target position,

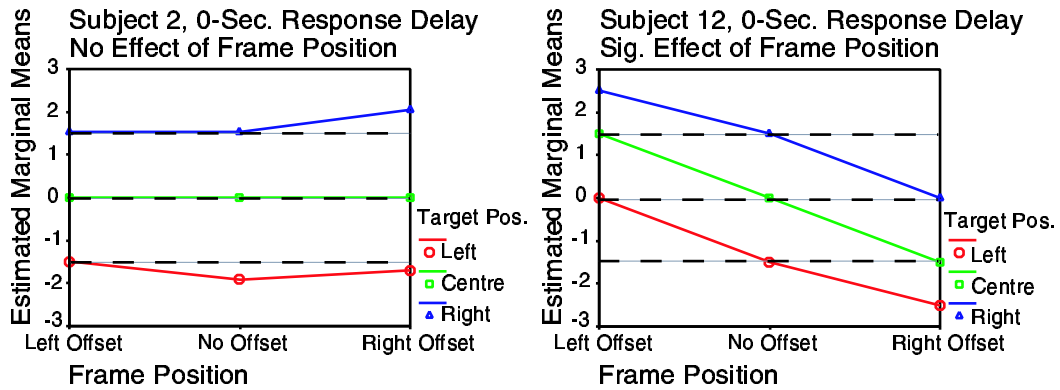


Figure 4.2: Estimated Marginal Means for Individual Cognitive Responses

while -1.5 degrees represents the left target position and $+1.5$ degrees represents the right target position. These marginal mean plots are representative of everyone that participated in the experiment; subject 2 is representative of those that showed no main effect of frame position, while subject 12 is representative of those that demonstrated an effect. For those subjects that demonstrated no effect, the marginal means indicate that variation in frame position did little to sway subject responses. However, for those subjects that demonstrated a main effect of frame position, the marginal means indicate that variation in frame position had a strong influence on subject responses. Thus, it appears that the verbal responses that subjects provided were quite consistent, and it is not likely that any observed effects are due to random noise in the collected data.

Five of the ten subjects that demonstrated a significant effect of frame position also showed marginally significant interaction effects. Although at first glance it might appear as though these higher-order effects should take precedence over any observable main effects, a closer examination of the data suggests that their presence does not imply a loss of main effect. Even the smallest F ratios for the observed main effects were four times larger than the F ratios of the corresponding interactions. For nearly all of the subjects, these F ratios were typically forty to fifty

times larger, meaning that these interaction effects are small comparatively to the observed main effects. Moreover, these interaction effects are not consistent across all of the subjects nor are they immediately interpretable in any reasonable way with respect to the effects that we are observing.

A descriptive analysis of subject performance during the cognitive report condition indicates that subjects' response times ranged from 1.8 seconds to 2.9 seconds. It was estimated that it took the experimenter exactly one second to enter subject responses into the experiment software application, indicating that the mean subject response time was approximately 2.2 seconds. There was no observed correlation between the presence or size of the induced Roelofs Effect and subject response times. Subjects consistently kept their heads steady throughout the condition. The largest measured head movements were 7.71 centimetres from nominal position, and 12.83 degrees from nominal orientation, where nominal position and orientation are determined by the first measured head movements for each particular subject.

A measurement of the number of practice trials presented to subjects provides some interesting results. Surprisingly, subjects seemed to have the most difficulty in completing the initial set of practice trials that included presentation of only targets. While some subjects were able to complete this initial set in the minimum ten trials, several of the subjects required twenty, or even thirty trials before they were able to meet the progression criteria. One subject required a total of 42 practice trials in the target-only presentations before advancement. The average subject required approximately twenty trials before advancement. Nevertheless, there were no observed correlations between the number of initial practice trials that were required and the presence or size of the induced Roelofs Effect in the subsequent experimental trials. A possible explanation for the diverse range in the number of trials required is that some subjects were dependent on visual cues in order to determine target positions while other subjects were able to use spatial coordinates

to determine target positions. Subject performance in the practice trials improved dramatically once the rectangular frames were added into the trial presentations. In the remaining two sets of target and non-offset practice trials, nearly all of the subjects advanced after only ten trial presentations. The largest number of trial presentations required for any one subject in these sets was fifteen trials, and the mean number of trial presentations was approximately eleven trials.

4.3 Open Loop Pointing Condition

An analysis of subject performance suggests that noticeable individual differences in performance characterize pointing without visual feedback in target acquisition tasks. Intuitively, we might expect that purely open loop pointing is likely to be highly unreliable as an interactive technique in virtual environments. Given the significant observed pointing errors across all of the subjects, this initially appears to be true to some degree. Nevertheless, the result of an analysis of the open loop pointing in this experiment suggests that even when pointing is performed without visual feedback, it is possible to aiming consistently over an extended period of time.

Figure 4.3 presents example scatterplots from several subjects that are overlaid on subject responses with the presented target positions. The figure highlights the four categories of major pointing errors that were observed during analysis. Two of the subjects were characterized as having significantly overaimed at target positions. Similarly, five of the subjects were observed to exhibit significant underaiming of the target positions. Three of the subjects demonstrated a tendency to aim too far to the left of the actual targets, while the remaining four subjects showed an opposite tendency to aim too far to the right of the actual targets. Thus, it appears that all of the subjects demonstrated significant pointing errors of one form or another. The pointing errors within each offset are also consistent with other studies of reaching and target acquisition; subjects demonstrated significantly smaller horizontal pointing variance when compared with their corresponding vertical pointing

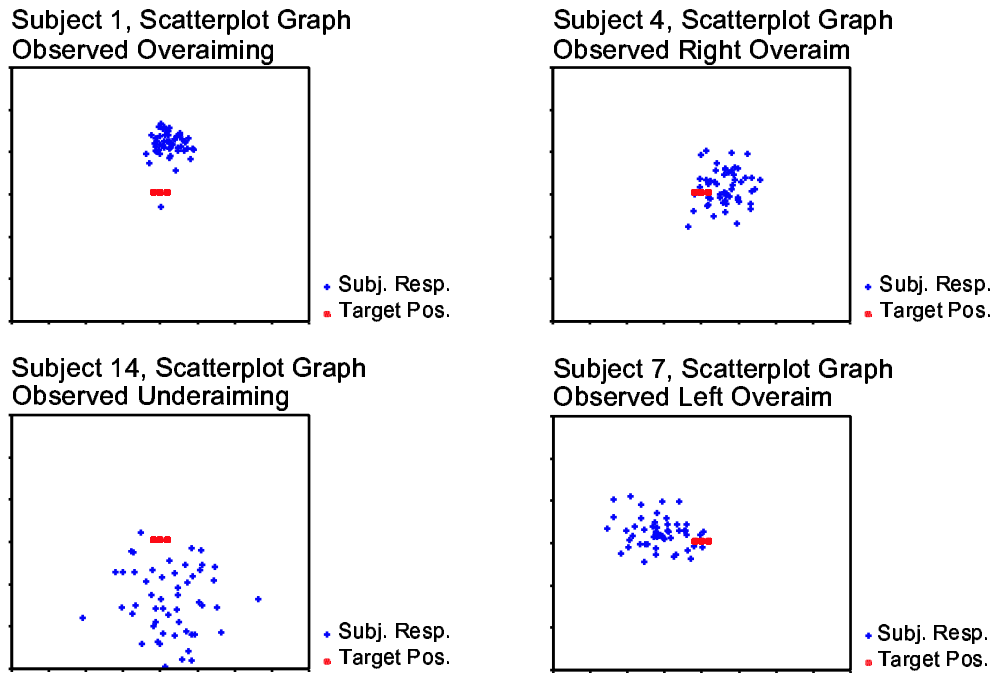


Figure 4.3: Scatterplots of Individual Open Loop Pointing Responses

variance (Jeannerod & Biguer, 1982). For this reason, and because we are looking at targets that only vary in position horizontally, these and subsequent pointing data are primarily analyzed along a horizontal axis of pointing.

All four of the subjects that demonstrated no induced Roelofs Effect in the cognitive report condition subsequently demonstrated no induced Roelofs Effect in the open loop pointing condition [$F(2, 18) < 2.401, p > 0.119$]. The mean magnitude of these subjects' responses was 1.2 degrees, with a standard error of 1.65 degrees. Given that these subjects appear to be consciously unaffected by the visual illusion, this result is not particularly surprising; such individuals may be able to successfully draw upon either of the cognitive or sensorimotor streams of visual perception equally well. It may also be that the cognitive representations of space for these particular subjects were relatively unaffected by the biases introduced by the induced Roelofs Effect, meaning that the presence of greater cognitive influence

had relatively little overall effect in terms of observed motor behaviour.

Six of these ten subjects failed to demonstrate a corresponding main effect of frame position [$F(2, 18) < 2.845, p > 0.084$] with immediate response delay in the open loop pointing condition, providing evidence for a dissociation of the cognitive and sensorimotor streams that is consistent with previous studies of the two visual systems hypothesis. The mean magnitude of these subjects' responses was 1.66 degrees, with a standard error of 1.70 degrees. Only one of the six subjects demonstrated no main effect of frame position in the immediate response delay condition. This subject demonstrated a highly significant main effect of frame position with a four second response delay [$F(2, 18) = 7.46, p < 0.004$], while the other five subjects continued to demonstrate no main effect of frame position. All ten of the subjects failed to show any significant interaction effects.

The remaining four subjects that demonstrated an induced Roelofs Effect in the cognitive report condition continued to demonstrate a main effect of frame position in the open loop pointing condition [$F(2, 18) > 5.072, p < 0.018$]. For these subjects, the mean magnitude of the effect was 5.80 degrees, with a standard error of 2.39 degrees. The presence of such subjects remains consistent with the results of past psychological studies of motor pointing and the two visual systems hypothesis, suggesting that these particular subjects may not have had the ability to retain the spatial information necessary in order to use their sensorimotor stream of visual perception to guide their aiming movements (Bridgeman et al., 1997).

Between all of the subjects, responses in the open loop pointing condition took significantly longer when compared against the estimated response times of the cognitive condition. In this motor condition, subject response times ranged from approximately 2.8 seconds up to approximately 11.5 seconds. The mean subject response time was calculated to be approximately 5.4 seconds. As with the measured head movements of the cognitive report condition, subjects did not move their heads a great deal throughout the open loop pointing condition. The largest recorded head

movements were nine centimetres from a nominal head position, and 15.60 degrees from a nominal head orientation.

4.4 Closed Loop Pointing Condition

Intuitively, we might expect that the presence of visual feedback in the closed loop pointing condition would provide consistently superior performance when compared against the characteristic performance of subjects in the open loop pointing condition. At a superficial level, this may be true because it is clear that the addition of a projected crosshair to subjects' pointing responses had a significant effect in reducing the overall aiming variance of subjects. Figure 4.4 provides an example of this in the form of a scatterplot comparison of open loop pointing against closed loop pointing for one participating subject in the experiment. Such a dramatic reduction in variance was representative of all subjects between the open loop and closed loop pointing conditions.

Nevertheless, a closer examination of the figure appears to suggest that there is at least some support for characterizing the remaining aiming variance in the closed loop pointing condition as nothing more than the aiming variance of the open loop pointing condition after it has been rescaled and translated. The difference in the represented coordinate axes of both scatterplots in the figure illustrates that once such a rescaling and translation occurs, open loop pointing variance is virtually indistinguishable from closed loop pointing variance. Knowing that we can account for the differences in precision between open loop and closed loop pointing is quite significant. If we can develop a VR system that can perform such a rescaling and translation for individual users, we may have at least one indication that the supposed advantages of visual feedback can actually be achieved in pointing without visual feedback. Thus, one of the more obvious reasons for using closed loop pointing vanishes and we have a stronger case for supporting open loop pointing.

This observation is further supported by an examination of the overall results

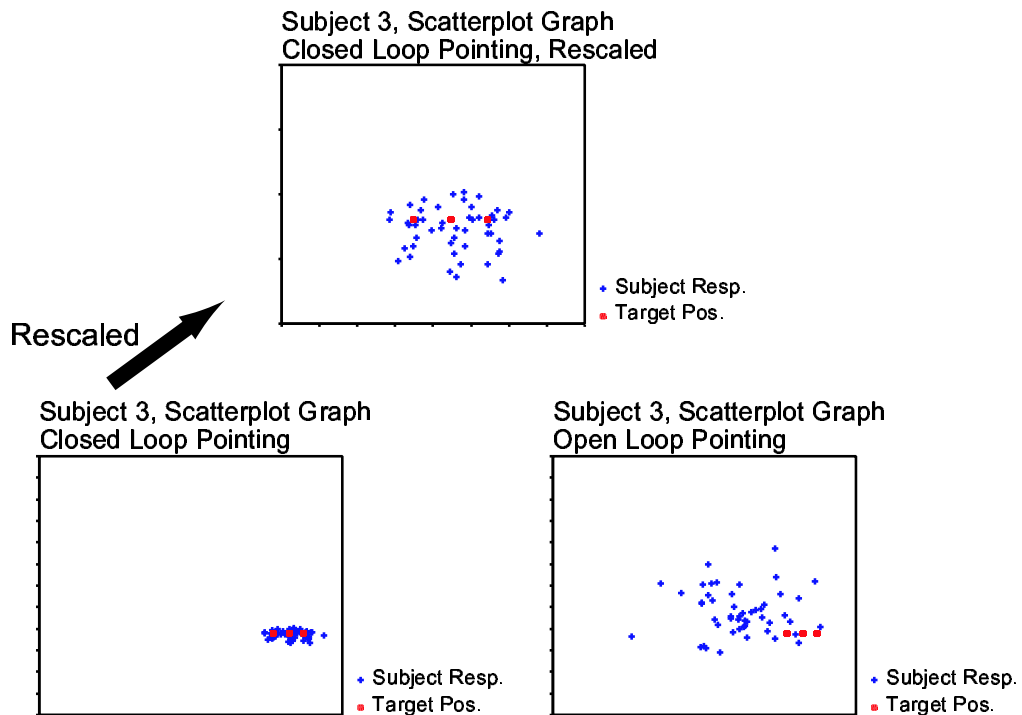


Figure 4.4: Scatterplots of Open Loop Versus Closed Loop Pointing Variance

that were provided by subjects in the closed loop pointing condition of the experiment. Of the six subjects that showed no induced Roelofs Effect in the open loop pointing condition, all of the six subjects demonstrated a corresponding main effect of frame position in the closed loop pointing condition [$F(2, 18) > 3.850, p < 0.05$], suggesting that the presence of visual feedback in their pointing responses actually caused them to incorrectly respond to the illusory positions of targets once again. None of the ten subjects showed any significant interaction effects. Thus, there appears to be some solid evidence that supports our experimental hypothesis concerning the relationship between the cognitive stream of visual perception and the presence of visual feedback in visually guided motor activity. For these six subjects, the mean magnitude of subject responses was 0.70 degrees, with a standard error of

0.38 degrees. The presence of the induced Roelofs Effect was consistent across both response delay conditions.

Of the four subjects that exhibited an induced Roelofs Effect in both the cognitive report and open loop pointing conditions, two of the subjects continued to show a significant main effect of frame position [$F(2, 18) > 3.966, p < 0.037$]. For these two subjects, the mean magnitude of the pointing responses was 0.72 degrees, with a standard error of 0.38 degrees. However, the remaining two subjects did not appear to show a correspondingly similar main effect [$F(2, 18) < 2.861, p > 0.083$]. The mean magnitude of pointing responses for these subjects was 1.46 degrees, with a standard error of 1.02 degrees. Although the results of these two subjects appear to be inconsistent with the results that we might expect of subjects that have consistently demonstrated induced Roelofs Effects in the other experimental conditions, an examination of the mean magnitude and standard error suggests that the lack of a significant effect might be attributable to noise or unexpected variance in these subjects' pointing responses. Given that these subjects are exhibiting no significant main effect of frame position, we should expect that the mean magnitude of their pointing responses would be lower than those of subjects who do exhibit such an effect. Thus, we have some reason for questioning the reliability of the closed loop responses of these two particular subjects.

Like the results of the open loop pointing condition, the four subjects that showed no induced Roelofs Effect in the cognitive report condition continued to show no induced Roelofs Effect in the closed loop pointing condition [$F(2, 18) < 1.311, p > 0.294$]. The mean magnitude of their responses was 0.28 degrees, with a standard error of 0.26 degrees. These subjects showed no evidence of any significant interaction effects. When compared to the response times of subjects in the open loop pointing conditions, the measured response times in the closed loop pointing conditions appear to be slightly better. Subject response times in this condition ranged from approximately 2.1 seconds to 5.2 seconds, with a mean response time

of 3.8 seconds. Once again, subjects did not make any large head movements during the closed loop pointing conditions. The largest measured head movement in this condition was a deviation of 8.19 centimetres from a measured nominal head position, and a deviation of 14.32 degrees from a measured nominal head orientation.

4.5 Closed Loop Pointing with Interactive Latency

The addition of interactive latency to the visual feedback of the projected crosshair during closed loop pointing provides some interesting insights into the effects of lag on visually guided motor actions. Consistent once again with the results of all of the previously reported conditions, these subjects continued to demonstrate no evidence of a significant main effect of frame position [$F(2, 18) < 3.226, p > 0.063$]. The mean magnitude of these subjects' responses was measured to be 0.37 degrees, with a corresponding standard error of 0.37 degrees. In addition, these subjects failed to show any significant interaction effects.

Interestingly, of the six subjects that failed to demonstrate a significant main effect of frame in the open loop pointing condition, three of the six subjects were assessed to also show no significant main effect of frame in this condition, suggesting that these subjects were able to maintain some level of open loop behaviour, as predicted by our experimental hypotheses. This contrasts with the previous closed loop condition that did not include any element of induced lag, where all six subjects demonstrated an induced Roelofs Effect: it appears that for at least some of the subjects, the presence of interactive latency created unreliable visual feedback that created a dissociation of the cognitive and sensorimotor streams of visual perception.

Figure 4.5 compares the estimated marginal mean plots between the closed loop pointing and closed loop pointing with interactive latency conditions for one of these three subjects. The other two subjects had marginal mean plots that were quite similar, with only minor differences that can be accounted for by the individual variations in pointing responses between the three subjects. Although it is

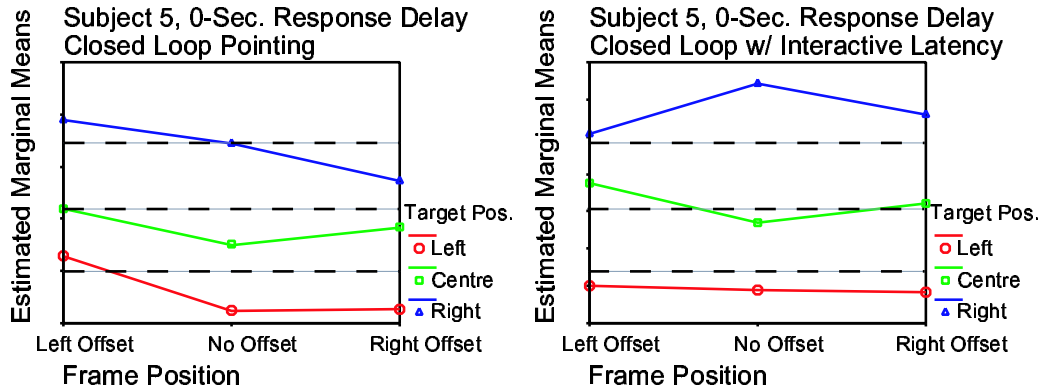


Figure 4.5: Comparing the Marginal Means of Both Closed Loop Conditions

not nearly as apparent as the differences between the cognitive report condition and the open loop condition, the comparison of the marginal mean plots appears to indicate that closed loop pointing with interactive latency generated fewer subject response mistakes with respect to frame position than the original closed loop pointing condition.

An interesting side observation to this is that only two of the three subjects failed to show any significant main effect of frame position across both response delay conditions. The remaining third subject exhibited a significant main effect of frame in the four second delay condition [$F(2, 18) = 6.479, p < 0.008$], with a mean magnitude of 1.06 degrees and a standard error of 0.60 degrees. Moreover, the subject that exhibited this behaviour was also the same subject that was only able to show no main effect of frame position with immediate response delay in the previous open loop pointing condition. Thus, there appears to be some additional support for the idea that this particular subject had a shorter spatial memory than the other subjects that failed to show an induced Roelofs Effect.

The four subjects that demonstrated an induced Roelofs Effect in both of the cognitive report and open loop pointing conditions continued to show a significant main effect of frame in this condition [$F(2, 18) > 4.294, p < 0.03$], with a mean

magnitude of 1.04 degrees and a standard error of 0.67 degrees. This result remains consistent with the explanation that these subjects did not have a sufficiently persistent spatial memory of target position that was accessible by the sensorimotor stream of visual perception. As such, these four subjects appeared to draw upon the allocentric memories that were provided by their cognitive perceptual streams in order to guide their aiming movements.

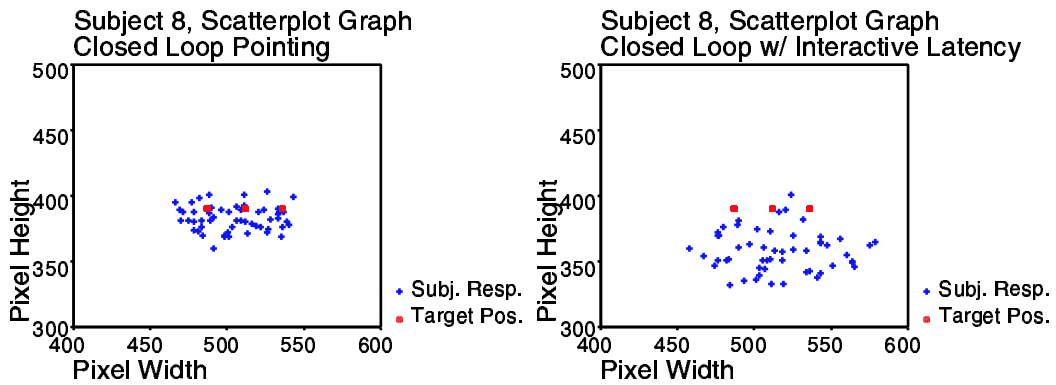


Figure 4.6: Comparing the Scatterplots of Both Closed Loop Conditions

It was clear that the presence of visual feedback, regardless of reliability, had an impact on the overall variance of pointing motions in this particular condition. Figure 4.6 presents a scatterplot comparison of one subject’s pointing responses between the previous closed loop pointing condition and this current condition. Generally, subjects demonstrated significantly reduced scatter variance in their pointing motions when compared relative to the open loop pointing condition, but this variance was not as tightly clustered as the variance reported in the previous closed loop pointing condition. Given the results of susceptibility to the induced Roelofs Effect, it may be that situations of closed loop pointing with interactive latency provide an intermediate level of performance between purely open loop and closed loop pointing.

An analysis of subject response time indicates that the amount of time to

make individual pointing movements in a closed loop pointing situation with sufficient interactive lag is quite similar to that required to make aiming movements in an open loop pointing situation. In the closed loop pointing with interactive latency condition, subject response time ranged from approximately 3.7 seconds to 7.5 seconds, with a mean subject response time of roughly 5.3 seconds. Consistent with the previously reported conditions, measured head movements suggest that subjects remained relatively still during the completion of this condition. The largest reported head deviations were 7.4 centimetres from a nominal head position, and 17.3 degrees from a nominal head orientation.

4.6 Observations and Subject Comments

In addition to the quantitative data that were collected from the participating subjects during the experiment, a number of qualitative observations were made. Subjects provided some interesting comments regarding their own personal experiences during the experiment. Given the results of the statistical analysis of subject performance, their comments provide a novel reflection of the difference between what we consciously perceive and how we act in virtual environments. In particular, subjects reported having much less confidence in their performance during the open loop pointing condition when compared to their performance during the closed loop pointing condition. Most of the subjects remarked that the lack of visual guidance made them feel as though their responses throughout the open loop pointing condition were random at best.

However, one subject also commented that an internal sense of proprioception appeared to provide some level of compensation for the lack of visual feedback. Although such insight into individual performance was likely the result of this particular subject's knowledge of human perception, the claim appears to have some validity. Jeannerod and Biguer (1982) recount analyses of the dynamic characteristics of reaching arm movements, demonstrating that there are systematic variations

of velocity during the trajectory of arm motion in open loop pointing situations, suggesting that when visual reaffirmation of target acquisition is not possible, another internal sense separate from vision attempts to compensate.

Such velocity variations were also quite apparent in the experimental condition of closed loop pointing with interactive latency. In this condition, subjects appeared to make initial ballistic aiming movements without reliance on projected crosshair position, suggesting that these initial movements are made in the same manner as open loop pointing. However, upon completion of these initial movements, slower movement “corrections” take place, suggesting that final movements are made under the guidance of the visual crosshair, and are thus made closed loop. As a consequence, it may be that in the presence of visual feedback there is an overriding dependence on visual reaffirmation, regardless of the correctness of the visual feedback. Thus, the presence of visual feedback may provide misleading reason to initiate intentional motor movements that cause unnecessary execution errors to occur.

At least one subject in the closed loop pointing with interactive latency condition was observed to attempt to compensate for the presence of lag by slowing her overall aiming movements. Although this subject’s pointing performance did not appear to differ dramatically from other subjects in this condition, it is an indication that the presence of lag may cause the development of different interactive strategies to compensate for the perceptual effects of lag. It also appears that inexperienced VR users are not necessarily capable of coping, nor expecting, the presence of lag in their overall interactive experience. In particular, upon initiation of the experiment trials in the closed loop pointing with interactive latency condition, one subject strongly asserted that the Fastrak stylus device was malfunctioning and suggested that “something” required repair before the remaining trials could proceed. Other subjects found that the lagged crosshair was more annoying than helpful, and it may be that some subjects simply ignored the presented crosshair in this condition,

which might have resulted in reversion to open loop motor behaviour. Subjects appeared to be unaware that the visual disconnection between pointing movements and projected crosshair were the result of lag, indicating that they felt that they had to “struggle” in order to get the crosshair to follow their pointing movements.

Observations made during the experiment and report from several subjects indicate that the confidence level of subjects in completing target acquisition tasks through pointing has at least some relationship to the perceived amount of hand jitter that we personally perceive. Studies of hand motion in laser pointer interactions suggest that the overall impact of hand jitter can mean absolute errors of approximately eight pixels on screen, which can be reduced to absolute errors between two and four pixels through filtering (Myers et al., 2002). Moreover, several subjects said that they felt at least somewhat lacking in confidence in their ability to correctly perform the required pointing tasks of the experiment because of the unsteadiness of their hands. Although perceived confidence did not appear to influence their overall ability to successfully complete the experiment, this does suggest that our level of confidence may influence our willingness to engage in direct manipulation tasks in VR applications for extended periods of time in more applied situations.

Because of the significant amount of time that was required to accomplish all of the experimental conditions, a particularly novel aspect of this experiment is that it allowed us to observe the overall comfort levels of users in large scale, immersive virtual environments. In particular, subjects frequently reported discomfort in wearing the stereo glasses over the course of the experiment. Some of the subjects observed that the stereo glasses were too big for them, suggesting that the glasses were at least a minor nuisance for them to wear during the experiment. Other subjects observed that the stereo glasses constricted their overall field of view. These subjects suggested that the presence of the stereo glasses could have affected their performance in the experiment. Additionally, subjects almost universally felt that the nature and duration of the prescribed experimental tasks contributed to some

feelings of tedium and boredom by the end of the experimental conditions. Although several of these subjects initially expressed an interest in the way in which they were interacting with the display environment, these effects of novelty eventually wore off over the course of the experiment. Explicit indications of fatigue and boredom following the experiment suggested that even the initial novelty of the virtual environment was insufficient to keep subjects engaged throughout the entire experiment.

4.7 Gender and Ordering Effects

Only three of the seven participating male subjects, but all seven female subjects in the experiment demonstrated an induced Roelofs Effect in the cognitive report condition, so there appears to be a gender difference in subject verbal performance. A simple *t*-test was run in order to confirm that this difference between genders was significant [$t = 2.83, p = 0.015$]. The subsequent motor pointing conditions did not show a similar gender difference, suggesting that only the allocentric, and not the egocentric, representations of space that are maintained by males and females may be somewhat different.

Although the induced Roelofs Effect in the cognitive report condition is a reliable phenomenon within a subject, given that there is a clear separation in response performance between those who did exhibit the effect and those who did not, it may be possible that the way in which the limited contextual cues of the surrounding virtual environment were processed led to a difference in the conscious visual perceptions between subjects. Cultural differences between the subjects may also explain the existence of this effect, although insufficient evidence was available to explore such an explanation.

The male subjects who did not exhibit the induced Roelofs Effect may have been more oriented toward the use of spatial cues in their surroundings, relative to those male subjects who did exhibit an induced Roelofs Effect. These particular

subjects might not have been inclined to use the visual perceptual cues provided by the rectangular frame in the target and frame presentations, meaning that their perception of space may have been inherently egocentric. This interpretation is augmented by the fact that all of the seven female subjects exhibited an induced Roelofs Effect. Previous studies have demonstrated that male subjects outperform female subjects in navigation tasks through virtual environments that have a particularly constrained field of view (Czerwinski, Tan, & Robertson, 2002), quite similar to the way in which males “outperformed” females in this experiment in their ability to avoid the perceptual influence of the induced Roelofs Effect. Thus, it is possible that the presence of surrounding contextual cues has an influence on the spatial impressions of surrounding environments in the cognitive stream of visual perception. This is consistent with the view of the cognitive stream as describing an allocentric representation of space, with a specific role of the cognitive stream being to take in visual information about details of shape and form, which are used to develop this representation.

There were no observable differences in the performance of subjects that experienced the open loop pointing condition prior to the closed loop pointing condition against those subjects that experienced the reverse order. As such, the counterbalancing of these two conditions saw no evidence of order effects. Specifically, there appears to be no connection between the order of these two conditions and the demonstration of significant main effects of frame position in any of the four conditions. For each of the four experimental conditions, some subjects from both counterbalanced groups exhibited a main effect of frame, suggesting that the presentation of one condition before the other is not a reasonable indicator of who is more likely to exhibit an induced Roelofs Effect. Moreover, the characteristic variances described by the scattering of pointing responses between subjects in each of the open loop and closed loop pointing conditions is not visibly different between those subjects who experienced one condition before the other. Thus, we conclude that

pointing performance did not appear to improve or worsen depending on presentation order.

4.8 Characterizing Subject Motor Performance

The experimental results that were derived from analyzing subject motor performance across the three pointing response modes indicate that there are differences in the kinds of pointing errors that are made between open loop and closed loop pointing conditions. In open loop pointing, errors do not manifest themselves in the form of misguided perception, as indicated by the presence of subjects that did not exhibit a significant main effect of frame position. However, errors appear to exist in the form of consistent, exaggerated aiming. Closed loop pointing exhibits the reverse pattern: errors of misguided perception occurred in subjects that demonstrated a significant main effect of frame position, while aiming variance appears to be small and well-controlled.

Our characterization of the induced Roelofs Effect solely as a main effect of frame position tells us little about the size or magnitude of its illusory influence. We are interested in determining the size of the actual effect because being able to separate the illusory effects from the aiming effects may further illuminate the distinction between both of these kinds of error, and it may also help us to better characterize the consistency of open loop versus closed loop pointing in virtual environments. We might suggest that an initial estimate of the effect size comes in the form of measuring the mean magnitude and standard error of subject responses. Because of the discrete nature of the cognitive report condition, such an estimation is actually quite reasonable; any subject verbal response directly corresponds to a discrete target position. As such, the magnitude of the induced Roelofs Effect can be directly determined from the verbal responses provided by the subjects.

However, across all of the motor pointing conditions, the relationship between the illusory effect and indicated target position is less clear. In these conditions, sub-

ject responses are measured in terms of ballistic aiming and the resulting projected screen position of the full pointing motion. Such measurements are highly unlikely to provide consistent measures of aiming toward a specific target position for even a small number of successive subject responses because there are likely to be individual differences in pointing between experiment trials. Since each participating subject was likely to have at least minor personal differences in execution of the pointing responses, this effect may be magnified between subjects. Thus, an initial examination of the mean magnitudes and standard errors that are associated with subject responses may not directly tell us as much about motor response performance as we might like.

In order to correctly characterize the magnitude of the induced Roelofs Effect, we decided to readjust the calculations for mean magnitude and subject responses in the motor response conditions by taking into account the fact that subject responses do not necessarily correspond to fixed target positions on the screen. Such a rescaling can be accomplished by first observing that pointing is still a consistent response across subjects, and that all of the subjects exhibited strong main effects of target position across all of the different pointing conditions. As such, we know that subjects consistently aimed more toward some particular location when targets were presented in a particular position. Thus, by calculating the mean positions pointed to, with respect to each presented target position in each pointing condition, we have an estimate that can be used to define *nominal target positions* that have a direct correspondence to each of the actual presented target positions during each of the pointing response conditions for each subject.

We can find the difference in magnitude between these nominal positions and the actual positions. Subtracting this magnitude difference out of the mean magnitudes of the subject responses, we are left with an estimate of the actual influence that is attributable to the induced Roelofs Effect. Table 4.2 and Table 4.3 present an evaluation of the mean magnitudes, both before and after they have been

	Unadjusted Mean Mag.	Adjustment	Adjusted Mean Mag.	Std. Error
Cognitive Report	1.67 degrees	0.00 degrees	1.67 degrees	0.26 degrees
Open Loop Pointing	5.80 degrees	5.10 degrees	0.70 degrees	2.39 degrees
Closed Loop Pointing	0.71 degrees	0.15 degrees	0.56 degrees	0.38 degrees
Closed Loop Pointing w/ Interactive Latency	1.00 degrees	0.37 degrees	0.63 degrees	0.63 degrees

Table 4.2: Mean Magnitudes of Subjects with Induced Roelofs Effect

	Unadjusted Mean Mag.	Adjustment	Adjusted Mean Mag.	Std. Error
Cognitive Report	0.25 degrees	0.00 degrees	0.25 degrees	0.41 degrees
Open Loop Pointing	1.43 degrees	1.70 degrees	0.23 degrees	1.68 degrees
Closed Loop Pointing	0.87 degrees	0.60 degrees	0.27 degrees	0.64 degrees
Closed Loop Pointing w/ Interactive Latency	0.36 degrees	0.25 degrees	0.11 degrees	0.35 degrees

Table 4.3: Mean Magnitudes of Subjects without Induced Roelofs Effect

rescaled in this fashion. These tables illustrate the impact that pointing accuracy had on evaluated effect size. Table 4.2 shows the mean magnitudes of subjects who exhibited induced Roelofs Effects, while Table 4.3 shows the mean magnitude of subjects who did not exhibit induced Roelofs Effects. With rescaled measures, the mean magnitudes of the induced Roelofs Effect fall within a difference of one target position, which is consistent with past studies of the two visual systems hypothesis. Thus, some of the excessively large mean magnitude values that were reported in the previous analyses may be largely accounted for by individual differences and pointing error.

Figure 4.1 shows that as we move from closed loop conditions to the open loop pointing condition, the number of subjects who exhibit the induced Roelofs Effect steadily decreases. The magnitude of the effect does not appear to change a great deal across conditions, suggesting that the true size of the induced Roelofs Effect in pointing during the experiment was approximately one half of one tar-

get position. The distinct difference in effect size between verbal response in the cognitive report condition and the pointing conditions appears to be quite large, although it is likely that this difference in effect size is largely the result of the forced five-choice response mechanism of the cognitive report condition versus the more continuous response mechanism of the other pointing conditions, where the effect sizes are more consistent.

4.9 An Alternative Analysis of Illusory Effects

Although our established criteria for determining the presence or absence of the induced Roelofs Effect for each subject up to this point has been a main effect of frame position with respect to provided responses at a $p \leq 0.05$ level of significance, there are other criteria that we can use to determine the presence or absence of the visual illusion. Performing an alternative analysis confers at least two benefits. First, an alternative analysis provides additional confirmation that the observed effects actually exist and are not due to experimental error. Second, an alternative analysis may permit us to look at the illusory effects in a different way, identifying performance characteristics that were not immediately visible through the initial inferential statistical tests of main effects.

In this section, we evaluate subject performance on a per-trial basis. In this type of evaluation, subjects are deemed to have exhibited responses with an induced Roelofs Effect if and only if fifty percent or greater of their responses to trials with frames that are offset to the left or right exhibit an induced Roelofs Effect and fewer than fifty percent of their responses to trials with frames that are not offset exhibit an induced Roelofs Effect. A particular trial is said to exhibit an induced Roelofs Effect if and only if the subject response is in a direction consistent with the definition of the induced Roelofs Effect. Specifically, if the frame is offset to the left, then we should expect that subject responses are to the right of the correct response and if the frame is offset to the right, then we should expect that subject responses

are to the left of the correct response. A trial with a non-offset frame position is stipulated to be incorrect if and only if it does not fall within some measured distance of the nominal target position that corresponds to the presented target position. We define nominal target positions to be roughly the same as those that were previously used to calculate the true magnitude of the induced Roelofs Effect for each subject. Additional “illusory” nominal target positions are calculated, which exist beyond the left and right presented targets at distances equal to the mean distance between the centre, left, and right nominal target positions.

In order to determine if a subject response is consistent with the definition of the induced Roelofs Effect, we check to see what frame position is indicated by the presented trial. If the frame is to the left, then we decide that a subject response is consistent if it is within some measured distance of the nominal target position that is to the right of the nominal target position that corresponds to the presented target position. Likewise, if the frame is to the right, then we decide that a subject response is consistent if it is within some measured distance of the nominal target position that is to the left of the nominal target position that corresponds to the presented target position. In order to provide a reasonable measure of distance, we measure consistency and correctness in terms of providing a response that is within one standard deviation of the subject’s expected response target.

Such a set of formal criteria capture the essence of the induced Roelofs Effect, while simultaneously enabling us to look at any illusory effects in terms of frequency of response instead of variance of response. In particular, this criteria allows us to look at the percentage of times an individual subject makes illusory response errors in each of the different response conditions. This is particularly interesting in the context of HCI research because we now have a measure that may be interpretable as a quantitative value to indicate the probability we expect populations of VR users to make perceptual and motor errors using the different interactive response techniques that were examined in our experiment.

Subject Number	Left Fr. Roelofs Effect Response Percentage (%)	Right Fr. Roelofs Effect Response Percentage (%)	Total Roelofs Effect Response Percentage (%)	Non-Offset Frame Incorrect Response Percentage (%)
1	5.6	5.6	5.6	5.6
2	11.1	5.6	8.3	16.7
3	0.0	11.1	5.6	11.1
4	5.6	27.8	16.7	5.6
5	100.0	100.0	100.0	0.0
6	77.8	72.2	75.0	22.2
7	100.0	100.0	100.0	0.0
8	100.0	94.4	97.2	5.6
9	94.4	100.0	97.2	0.0
10	88.9	83.3	86.1	0.0
11	100.0	100.0	100.0	0.0
12	83.3	83.3	83.3	11.1
13	88.9	88.9	88.9	16.7
14	88.9	83.3	86.1	22.2
Means	67.5	68.3	67.9	8.3

Table 4.4: Subject Response Percentages in Cognitive Reporting

Table 4.4 presents a summary of the effective response frequency of the induced Roelofs Effect per subject in the cognitive report condition. The listed percentages reflect the frequency with which subjects provided verbal responses that could be classified as exhibiting an induced Roelofs Effect. The rows that are highlighted in bold indicate subjects that met the alternate criteria for exhibiting an illusory effect overall. These values have a very strong and consistent relationship with the main effects that were observed by the two-way ANOVAs, as indicated by the second column of Table 4.1. For those subjects that did not exhibit an induced Roelofs Effect, subject responses showed consistently small occurrence frequencies of the effect. In contrast, those subjects that did exhibit an induced Roelofs Effect showed consistently large occurrence frequencies of the effect. Thus, we have further evidence to support our interpretation of the distinct subject populations that are observed as a result of the cognitive effect of this particular visual illusion. The number of non-offset trials where subjects provided incorrect responses was also correspondingly small, below 22.2 percent in all cases, indicating that the presence

of the induced Roelofs Effect was reliable and was not due to the presence of overall random variations in subject responses.

Subject Number	Left Fr. Roelofs Effect Response Percentage (%)	Right Fr. Roelofs Effect Response Percentage (%)	Total Roelofs Effect Response Percentage (%)	Non-Offset Frame Incorrect Response Percentage (%)
1	27.8	33.3	30.6	22.2
2	27.8	22.2	25.0	22.2
3	33.3	33.3	33.3	44.4
4	38.9	27.8	33.3	66.7
5	38.9	27.8	33.3	11.1
6	33.3	33.3	33.3	11.1
7	27.8	27.8	27.8	22.2
8	44.4	33.3	38.9	22.2
9	44.4	44.4	44.4	11.1
10	50.0	44.4	47.2	5.6
11	33.3	44.4	38.9	22.2
12	44.4	50.0	47.2	11.1
13	27.8	33.3	30.6	22.2
14	38.9	38.9	38.9	11.1
Means	36.5	35.3	35.9	21.8

Table 4.5: Subject Response Percentages in Open Loop Pointing

Table 4.5 presents a summary of the effective response frequency of the induced Roelofs Effect per subject in the open loop pointing condition. Overall, the alternate criteria is strongly consistent with the lack of significant main effects of frame position that were established by the previous two-way ANOVAs in open loop pointing. Specifically, all of the individuals who were shown not to have any significant main effects also failed to meet the alternate criteria established here. However, our criteria does appear to indicate that the lack of the induced Roelofs Effect in open loop pointing is even stronger than suggested by our initial two-way ANOVAs; none of the subjects were shown to meet the alternate criteria for overall exhibition of the induced Roelofs Effect, as indicated by the third column of Table 4.1. A simple interpretation of this discrepancy may be the difference in which presence of the visual illusion is observed. In the previous two-way ANOVAs, this presence was determined by relative variance from established response means, while in the

alternate criteria presence was determined by absolute numbers of trials. Thus, the alternate criteria might be considered to be more “careful” in its categorization of subjects.

Subject Number	Left Fr. Roelofs Effect Response Percentage (%)	Right Fr. Roelofs Effect Response Percentage (%)	Total Roelofs Effect Response Percentage (%)	Non-Offset Frame Incorrect Response Percentage (%)
1	27.8	27.8	27.8	5.6
2	16.7	11.1	13.9	11.1
3	27.8	27.8	27.8	22.2
4	33.3	27.8	30.6	5.6
5	33.3	50.0	41.7	5.6
6	22.2	16.7	19.4	5.6
7	55.6	55.6	55.6	0.0
8	55.6	72.2	63.9	0.0
9	88.9	66.7	77.8	11.1
10	88.9	77.8	83.3	0.0
11	61.1	88.9	75.0	5.6
12	61.1	61.1	61.1	0.0
13	61.1	44.4	52.8	0.0
14	44.4	44.4	44.4	5.6
Means	48.4	48.0	48.2	5.6

Table 4.6: Subject Response Percentages in Closed Loop Pointing

Table 4.6 presents a summary of the effective response frequency of the induced Roelofs Effect per subject in the closed loop pointing condition. Once again, the results of the alternate criteria are consistent with the presence of significant main effects that were established by the previous two-way ANOVAs in closed loop pointing, as indicated by the fourth column of Table 4.1. A comparison of overall subject performance between open loop and closed loop pointing suggests that closed loop pointing is more susceptible to the perceptual influence of the induced Roelofs Effect than open loop pointing seems to be. Although the present analysis tells us very little about differences in repeatability (precision) across open loop and closed loop pointing, it does provide additional quantitative confirmation that open loop pointing may actually be more accurate than closed loop pointing. Thus, the alternate analysis provides more evidence to suggest that open loop pointing may

be a viable substitute for pointing interactions that provide visual feedback.

Subject Number	Left Fr. Roelofs Effect Response Percentage (%)	Right Fr. Roelofs Effect Response Percentage (%)	Total Roelofs Effect Response Percentage (%)	Non-Offset Frame Incorrect Response Percentage (%)
1	33.3	22.2	27.8	5.6
2	22.2	22.2	22.2	33.3
3	38.9	38.9	38.9	44.4
4	33.3	38.9	36.1	5.6
5	27.8	22.2	25.0	11.1
6	33.3	33.3	33.3	11.1
7	27.8	27.8	27.8	5.6
8	83.3	88.9	86.1	16.7
9	44.4	38.9	41.7	11.1
10	94.4	94.4	94.4	38.9
11	38.9	38.9	38.9	5.6
12	77.8	77.8	77.8	0.0
13	38.9	33.3	36.1	11.1
14	44.4	44.4	44.4	22.2
Means	45.6	44.4	45.0	15.9

Table 4.7: Subject Response Percentages in Closed Loop Pointing with Lag

Table 4.7 presents a summary of the effective response frequency of the induced Roelofs Effect per subject in the closed loop pointing with interactive latency condition. Consistent with the application of the alternate criteria to all of the other experimental conditions, the results of applying the present criteria appear to provide additional evidence that supports the observed significant main effects of frame position that were derived from the initial two-way ANOVAs, as indicated by the fifth column of Table 4.1. In particular, this analysis quantitatively distinguishes those subjects who exhibited an induced Roelofs Effect from those who did not. The three indicated subjects who exhibited the effects of the visual illusion appeared to exhibit it quite reliably throughout all of their trials with particularly high response frequencies when compared to all of the other subjects in the condition. Although there is insufficient data available to make a firm interpretation as to why this is the case, these data are consistent with the interpretation that these subjects resorted to open loop pointing behaviour in the presence of unreliable visual feedback.

Chapter 5

Discussion and Interpretation

The results of the current experiment lend themselves to some very interesting interpretations that apply to the development of immersive virtual environments. Moreover, these results furnish the evidence that we require in order to provide reasonable support for the verification of our two-visual systems model of perception and action in virtual reality. In particular, the results provide a good demonstration of the potential power of the developed model in helping us understand the perceptual and motor behaviours of users in VR. The experimental results, in conjunction with the two visual systems model, provide a solid foundation for explanation and discussion. Specifically, the model can be used to explain a variety of previously reported user phenomena in the literature, and it also provides a cogent explanation of the ancilliary effects that were observed in the statistical analysis of subject performance in our experiment.

5.1 Two Visual Systems Phenomena in VR

The presence of strong evidence to suggest that we can create an induced Roelofs Effect that creates a misperception of target position in virtual environments is particularly important when we consider the design of large scale, immersive environments that must cope with providing users with all of the contextual cues that

are necessary for promoting correct, efficient behaviour in a consistent and reliable manner. At first glance, the environmental tasks and conditions imposed on participating subjects during the current experiment may not appear to be entirely reflective of typical situations that employ VR applications, even for those applications that involve repeated target acquisition. This is not entirely unreasonable, given that the experiment was designed to be executed in a variation-free environment that was supposed to maximize the effects of any expected phenomena. However, upon closer inspection, there are at least a couple of reasons to suggest that variations of the induced Roelofs Effect as presented in the present experiment may commonly occur in VR applications.

First, the presence of contextual cues is a particularly influential factor in our perception of orientation in large scale display environments. In the present experiment, this is a possible explanation for the difficulty that subjects had in progressing from target-only practice trials to target and non-offset frame practice trials. It also helps explain why male subjects appeared to be less susceptible to the induced Roelofs Effect than female subjects. Because large scale environments in the absence of display stimuli are relatively dark and homogeneous physical configurations that provide relatively few contextual cues, users of such immersive displays are dependent on the rigid contextual cues that are provided by display presentations in order to establish a suitable spatial frame of reference. Since our visual perception is not particularly selective about the kind or quality of the contextual cues that are provided in order to establish such a frame of reference, developers of VR applications should be aware of the kinds of effects that misleading contextual cues may have on user perception and action. For example, if we find ourselves in a large-scale graphical user interface that involves the presentation of rectangular window frames as a major organizational component, an induced Roelofs Effect may come in the form of misperceiving and misdirecting motor manipulations toward the locations of graphical widgets inside these window frames.

Second, it is important to recognize that the general perceptual effects described by the Roelofs Effect comes in many forms, and that the particular form that was used in our experiment is just one example of what an induced Roelofs Effect may look like. In a virtual environment that is significantly decontextualized, the induced Roelofs Effect may manifest itself in the form of the rectangular frame and circular target that we have used, or it may manifest itself in the form of rendered flashing stimuli that have similar potential to be misperceived. Because we have been able to demonstrate the ability of one particular form of the Roelofs Effect to have a significant perceptual influence, and because we can attribute this perceptual influence to the overall Roelofs Effect, it is not unreasonable to suggest that other “induced” forms of the Roelofs Effect might exhibit the same kinds of perceptual effects. These induced forms of the Roelofs Effect may be present in VR direct manipulation tasks that are not purely the simple acquisition of small targets in a large scale display.

Generally, the validation of our two visual systems model implies that any of the other visual illusions that have been used to demonstrate a dissociation of the cognitive and sensorimotor streams of visual perception in previous studies are also potential influencing phenomena in virtual environments. As such, visual illusions that induce unconscious alterations to internal perceptual processing, such as those introduced by invocation of saccadic suppression, induced motion effects, or size-contrast illusions, are possible sources of perceptual error. With the inclusion of an even wider variety of potentially deceiving visual stimuli, the proposed two visual systems model can be seen to have a greater range of applicability. The wide variety of physical characteristics, possible alternative configurations, and perceptual influences of these visual illusions means that the two visual systems model could be applicable to any situation that relates to the development of user interaction techniques that involves direct motor manipulation within a virtual environment.

5.2 Developing VR with the Two Visual Systems Model

The consequences of our two visual systems model of perception and action are only useful to us if we can convert the predictions that are generated by the model into design principles that will help us develop VR systems in the future. Considering the results of the present experiment, we observe that significant individual differences characterize the interaction styles of each subject. Whether these individual differences result in unique pointing errors, different reaction times, or susceptibility to perceptually-influencing visual stimuli, one of the most important ideas that comes from our model of perception and action is that no matter how well designed a user interface is, a simple interface design cannot be expected to cope with all of the minor variations that differentiate one user from another. It is at least plausible to argue that the user interfaces provided by many VR applications make no dynamic affordances for users within a specific application. They merely assume that an interface that is based on direct manipulation techniques is all that is required in order to make virtual environments usable.

We should be aware of the particular instances in which it is especially appropriate to apply our two visual systems model of perception and action toward the development of VR applications. Outside of the more universally applicable claims that are described by the two visual systems model, there may be specific instances in which it makes more sense to pay attention to the two visual systems model in order to provide theoretical insight into the development of VR systems. For example, while it may not necessarily make sense to base the entire design of a VR text editing application on the two visual systems model, it may be prudent to seriously consider the predictions made by the two visual systems model in rich sensory environments that involve complex interactions with fast-moving visual stimuli. Such environments are reminiscent of those that involve heavy, simultaneous use of multiple perceptual modalities, cognition, and various motor functions. These kinds of environments are most often found in vehicle design, flight simulation, or more

generally in any of a number of different domains where large-scale, real-time systems are employed in order to improve the operational efficiency that is dictated by a particular situation. For these kinds of virtual environments, the two visual systems model tells us that we should always be aware of the stress and subsequent effects that we place on users when we provide them with visual information that they must use to accomplish a complex task.

We may do well to consider the idea of developing user interfaces that not only present users with robust environments, but interfaces that also adapt themselves to meet the requirements of particular tasks and particular users. For complex virtual environments, this might mean developing user interfaces that use the predictive power of our two visual systems model in order to anticipate the reactions of users and to respond accordingly. In anticipation of user behaviour, such VR user interfaces might be able to identify potentially influential visual illusions and compensate for any unconscious errors of execution that might be generated from their presentation. Furthermore, these user interfaces might be capable of predicting users' motor movements and visually guided actions, compensating for minor incongruencies in human motor behaviour that impact user performance, such as hand jitter and the effects of overaiming in environments that require aiming and target acquisition. Such adaptive designs provide a level of flexibility that permits users to make mistakes, whether these mistakes are unconscious or uncontrollable, while still correctly resulting in the intentions of users being carried out by the VR system.

In the absence of the capacity to develop such ideal interface designs, our two visual systems model of perception and action emphasizes the importance of minimizing the presence of visual phenomena that might cause us to make perceptual errors. It also provides us with some critical motivation for informing developers that the design of virtual environments should prevent users from getting into situations where they must overcome the limitations of the manipulations provided

by the interface in order to achieve what is actually desired from the system. One particularly striking example of this is provided by the results of our present experiment, in the performance differences between verbal and motor report as response mechanisms for direct acquisition of targets.

It has been anecdotally remarked on various occasions that voice recognition may make an ideal mechanism for manipulation in computing environments once the technology for recognition has improved to the point that it is technically feasible and reliable because it is a quick and natural response mechanism for users. However, our results indicate that this is not necessarily true: responses that were influenced by cognitive visual processing led subjects to make mistakes in acquiring correct target positions verbally that were not similarly present under open loop pointing conditions. If the tasks specified by the experiment had been put into the context of a real situation, such as in the acquisition of targets in a large scale air traffic control system, the consequences for incorrectly assessing and verbally reporting target position might have been severe. Even if users are aware of the potential pitfalls of using such a response mechanism, there may be no easy way to train users to avoid such dangers.

Our two visual systems model tells us that we should carefully discriminate between situations where particular interaction styles are appropriate and where they are not appropriate. By using the two visual systems model to predict user behaviour in virtual environments, we may be provided with some insight as to whether or not a particular interactive technique is safe and reliable for a particular situation. We may also obtain clues as to what kinds of techniques are most applicable in specific VR situations, and where some of the performance boundaries for these techniques may lie.

A curious aspect of our model is the psychological framework upon which it is based. The study of human behaviour is vast and wide-reaching, and it is simply impossible to encompass everything that could possibly pertain to VR in a single

theory or model. Thus, we should not look upon our current model of perception and action as being the only aspect of human behaviour that should concern us in virtual environments. Instead, we should view this model as a single module in a large collection of modules regarding human behaviour in virtual environments. The two visual systems hypothesis provides us with a reliable guide for predicting human behaviour and the relationship between visual perception and action. Although many such models have yet to be fully formalized and developed, there are presumably other modules that could be used as reliable guides to other aspects of human behaviour. These models could conceivably cover areas of human behaviour including other relationships between perception and action as they apply to other sensory modalities, such as audition. The important idea to recognize here is that we should be aware of as many of these “modules” as possible, applying them wherever we can in order to improve our design methodologies for virtual reality.

5.3 Why Consider Open Loop Manipulation?

In assessing the performance differences of interactive pointing under varying levels of visual feedback, the current experiment appears to suggest that open loop pointing may be preferable to closed loop pointing in some VR situations, provided that there is a way for us to control for the motor error inherent in making skilled movements without the aid of visual feedback. The experimental results and the two visual systems model suggest that open loop motor manipulations of any kind may be preferable to conventional closed loop manipulations. These specific findings may not necessarily fit our intuitive model of the world. Consistent with many of the subjects’ beliefs about their open loop pointing performance during the present experiment, it may be difficult for us to accept the possibility that our motor actions may actually be more reliable when we are provided with less conscious information about what we are doing. Thus, we may find ourselves asking why we would even want to consider developing open loop interactive techniques for virtual

environments.

Card, Moran and Newell (1983) present a complete model of human information processing that includes estimations of temporal motor performance. Their model suggests that closed loop motor movements are the result of time spent performing perceptual processing, cognitive processing, and motor processing. In their estimation, a single closed loop motor “micromovement” should take approximately 240 milliseconds. However, the actual motor processing only represents a small fraction of the overall processing time; it is claimed that a single open loop motor micromovement can be executed on the order of 70 milliseconds. If their estimations of motor performance are accurate, execution of motor movement under purely open loop conditions should be faster than motor movements under closed loop conditions. While Card, Moran and Newell provide no indication of how we might account for such a performance increase, our two visual systems model provides an adequate explanation: there is a dedicated sensorimotor stream of visual processing that directs spatial motor movements more effectively than a corresponding cognitive stream of visual processing.

In addition to potential temporal performance gains, open loop motor manipulations implicitly reinforce the importance of actually completing motor tasks in virtual environments instead of concerning ourselves with the specific details of how to accomplish specific motor tasks. In VR applications that provide visual feedback, the two visual systems model indicates that there is a relationship between what we consciously perceive and how we act. Thus, we have reason to believe that the presence of visual feedback may cause users to spend unnecessary conscious effort in assessing the most effective method for accomplishing a motor movement, despite the fact that we already have a stream of visual processing that can conduct such an assessment at an unconscious level. Open loop motor manipulations are not likely to be penetrable to such unnecessary cognitive processing because the only visual information that we receive is information about what is in the environment, and

not how this environment is best manipulated by the user.

This leads directly into the idea of promoting open loop motor activities as one method for reducing the amount of cognitive effort that is required for users to assimilate the enormous quantities of information that are present in sensory-rich virtual environments. Because motor manipulations without visual feedback engage users to think less about what kinds of motor actions they are performing and to spend more time performing these actions, reduced strain is placed on users who must cope with vast amounts of sensory information. In potentially stressful situations, where motor movements are used in the context of skilled vehicle operations or time-limited responses, the ability to reduce the cognitive load of users is always an advantage. By reducing the amount of emphasis that must be placed on performing motor actions, we offload cognitive effort that can be used elsewhere in other aspects of the virtual interaction, generating an environment that is less stressful overall for users.

Open loop interactions in virtual environments may also help reduce the negative performance influences of interactive latency. The perceptual discord induced by interactive lag directly results in the discomfort of users in VR applications. Because system-induced lag may be unavoidable in many VR situations, we might reduce the effects of lag in at least three ways when we employ open loop motor activity. First, the removal of lagged visual feedback reduces the number of visibly lagged objects in the environment, potentially making it somewhat easier for users to cope with the perceptual effects of lag. Second, if users are permitted to perform motor actions in an open loop fashion, they may become more accustomed to the lag, improving performance and the overall user experience. This may be most noticeable in instances where lag is uniformly present in all components of the VR system: users may be aware of a uniform delay between their actions and the corresponding response from the system, but users may still be able to feel as though they can react normally. Third, the presence of visual feedback may ac-

tually be a contributing factor to the ill effects of interactive latency, especially if the visual feedback provided by the environment is particularly elaborate. Thus, by eliminating visual feedback in relation to motor responses, we may be able to reduce the overall amount of lag that is induced by the system.

Our current experiment is not the only formal evidence that exists to suggest that open loop manipulations may be a reasonable alternative to closed loop interaction in many situations. Some studies into interactive techniques both with and without visual feedback indicate that there are many instances where the addition of visual feedback does not necessarily result in significant performance improvements, while other studies indicate that the presence of visual feedback appears to require tighter control of system response time than may be achievable in particular virtual environments (Poupyrev et al., 1998; Watson, Walker, Ribarsky, & Spaulding, 1998). As such, the development of open loop interactive techniques may be a method for circumventing such problems that are associated with the presence of visual feedback.

Interestingly, such studies fail to provide a substantive explanation for the inability of visual feedback to improve user performance. These studies often do not attribute performance decreases to the presence of visual feedback, but to possible problems with the specific interaction styles that are being evaluated. While such explanations are certainly possible, our two visual systems model offers an alternative, possibly more attractive, explanation: visual feedback may unduly induce poor cognitive judgement through the cognitive stream of visual perception, thereby negatively influencing the performance of users that are engaged in certain motor manipulation situations in virtual environments.

Chapter 6

Conclusion

Through an exploration of virtual reality and experimental psychology, this thesis has demonstrated how the methodologies and techniques of two distinct domains can come together in order to produce fruitful results in the interdisciplinary domain of human-computer interaction. Our examination of the two visual systems hypothesis has a formal model of perception and action in virtual environments. Through the design, implementation, and analysis of an experiment in virtual reality, we have touched on some of the major issues surrounding the development of virtual environments, some of the key problems that face contemporary VR applications, and how a reflective approach to the development of VR systems can help us better understand where current systems need improvement, and how to avoid common pitfalls in the future.

Our two visual systems model of perception and action teaches us to recognize the inherent complexity of the human visual system and its tight relationship with the human motor system. Although our intuition might suggest that the relationship between visual perception and motor action is relatively straightforward, the vast amount of evidence that has been collected over many years in experimental psychology tells a very different story. The presence of parallel cognitive and sensorimotor streams of visual perception has many possible implications for the design

of large scale, immersive virtual environments, where visual context is limited and direct manipulation of the environment is accomplished through motor movement. Their presence also affects the development of safety-critical systems that include elements of virtual reality in their design; believing that objects in our surrounding environment are in one location, and acting upon these objects at a different location could be dangerous, indicating to us that we always need to be aware of the psychological underpinnings that integrate our mental and physiological mechanisms together.

The presented experiment has not only provided us with some credible evidence to support the two visual systems model in VR, but has also presented us with a close look at the nature of direct manipulation in immersive virtual environments. The results of the experiment clearly demonstrate that the behavioural phenomena that were first established in previous psychological studies of the two visual systems hypothesis directly apply to virtual reality. These results also show how the illustrative power of the two visual systems model can be used in order to effectively predict the verbal and motor behaviours of participating subjects. Through the experiment, we have seen how user performance changes under different open loop and closed loop response situations. In particular, we have seen how the two visual systems model accurately predicted the superiority of open loop pointing performance over that of closed loop pointing performance, as shown by the results of the experiment.

Although this was not its primary purpose, the experiment also provided us with a good opportunity to study users in a large-scale virtual environment for extended periods of time. Participating subjects in the experiment provided some insight into their own personal experiences, which are quite helpful in ascertaining the advantages and corresponding disadvantages that are conferred by using contemporary VR technology. In particular, the subjects in the experiment were quite eager to remark how their novel sentiments toward VR quickly changed to some level of discomfort because of the non-ergonomic focus of the VR equipment that

was used. Several subjects complained about the size and bulkiness of the equipment, which were mostly directed toward the size and weight of the stereo glasses that they were required to wear during their sessions. Thus, it seems reasonable to suggest that one of the more prominent weaknesses of virtual reality in its current form remains its focus on technology, instead of the users who use this technology.

Even though VR technology initially seems quite impressive to the inexperienced user, such novelty alone cannot be expected to compensate for the drawbacks of many areas of VR technology. It is for this reason that it is fortunate that VR hardware and software are maturing to the point that usability is now becoming a key focus in the development of virtual environments and VR-style applications. For example, the relative bulkiness of the active stereo glasses may not be an issue in the future, as lightweight and more suitably-sized passive stereo glasses, as well as other kinds of autostereoscopic displays make their way into commercial production. We remain hopeful that a similar transition toward more usable software follows the trend toward more usable hardware in VR. By constantly improving our understanding of interactive user interface design in immersive environments and by further developing our knowledge of direct manipulation techniques in large scale systems, we can expect that future VR applications will not be encumbered by the many problems that appear to plague currently existing applications.

The path to much of this continuing innovation in VR awaits future research in HCI. In relation to the two visual systems model of perception and action, future research could come in the form of progressive studies of users in virtual environments. Future experiments might look into other kinds of visual stimuli that can dissociate the two streams of visual perception from one another. They might also look at different kinds of VR tasks, perhaps entirely unrelated to the target acquisition tasks that were prescribed by the current experiment. More importantly, these experiments may also use the current experiment as a guide for improving the experimental methodologies of future studies. For example, the present experiment

used a discrete measurement in order to judge the effects of illusory influence with a verbal response mechanism. Future studies may wish to pursue the use of a more continuous measurement in order to look at the effect size of visual illusions. Other such studies may even wish to substitute the verbal report mechanism with another form of cognitive report in order to determine if there are any significant differences.

Because the current experiment was based on psychological methodologies for examining the two visual systems hypothesis, the issue of studying the two visual systems model of perception and action in a more applied environment still remains open. It might be particularly enlightening to see what the two visual systems model can tell us about more situations where direct motor manipulation plays an important role. It would be equally enlightening to see how the two visual systems model holds up in such situations. As a related future topic for research, it might be interesting to measure the perceptual limitations of visual illusions in different kinds of display environments. Common representations of visual illusions are usually limited to monocular presentations. It would be interesting to see if these visual illusions are equally effective in fish tank virtual environments, head mounted displays, and in display environments that include some mechanism for head-coupled perspective projections and stereoscopic imagery. In relation to more static illusions, such as the induced Roelofs Effect used in the current experiment, future studies may look into measuring the just-noticeable-difference values for offset frames and targets.

As an individual component of a complete model of cognitive behaviour in virtual environments, it may be that the most significant contribution that the two visual systems model makes is that it motivates us to look into other sources of related behavioural theory that are described in experimental psychology and in other such disciplines. In such behavioural theory, we may find useful information about how we might go about improving the development practices of VR applications that are yet to be developed. These kinds of behavioural theory may form the basis

for other models that predict and explain user behaviour in virtual reality, helping us move closer toward the goal of developing a complete set of models that we can draw upon during the development of virtual environments wherever and whenever it is fitting to do so. In this way, through constructs such as the two visual systems model of perception and action and other similar models, virtual reality may continue to make further impressions on the way that we look at computing and technology in the future.

References

- Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, 5(6), 679–685.
- Ardizzone, E., Chella, A., & Pirrone, R. (2000). An architecture for automatic gesture analysis. *Proceedings of AVI 2000*, 205–210.
- Aukstakalnis, S., & Blatner, D. (1992). *Silicon mirage: The art and science of virtual reality*. Berkeley: Peachpit Press.
- Azuma, R. T. (1997). Correcting for dynamic error. *ACM SIGGRAPH '97 Course Notes*.
- Azuma, R. T., & Bishop, G. (1994). Improving static and dynamic registration in a see through hmd. *Proceedings of SIGGRAPH '94*, 197–204.
- Bajura, M., Fuchs, H., & Ohbuchi, R. (1992). Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *Proceedings of SIGGRAPH '92*, 203–210.
- Bolt, R. (1980). "put-that-there": Voice and gesture at the graphics interface. *Computer Graphics*, 14(3), 262–270.
- Bowman, D. A., Johnson, D. B., & Hodges, L. F. (1999). Testbed evaluation of virtual environment interaction techniques. *Proceedings of VRST '99*, 26–33.
- Bridgeman, B. (2000). A sensorimotor map of visual space. *Cognitive Science 2000*.
- Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, 15, 719–722.
- Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception and Psychophysics*, 29(4), 336–342.

- Bridgeman, B., Lewis, S., Heit, G., & Nagle, M. (1979). Relation between cognitive and motor-oriented systems of visual position perception. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 692-700.
- Bridgeman, B., Peery, S., & Anand, S. (1997). Interaction of cognitive and sensorimotor maps of visual space. *Perception and Psychophysics*, 59(3), 456-469.
- Brown, J. (1996). *Methodologies for the creation of interactive software* (Technical Report No. CS-TR-96/1). P.O. Box 600, Wellington, New Zealand: Victoria University of Wellington.
- Browning, D., Cruz-Neira, C., Sandin, D., & DeFanti, T. (1993). The cave automatic virtual environment: Projection-based virtual environments and disability. *Proceedings of the First Annual International Conference, Virtual Reality and People with Disabilities*.
- Bruce, V., & Green, P. (1990). *Visual perception: Physiology, psychology, and ecology*. East Sussex, United Kingdom: Lawrence Erlbaum.
- Busettoni, C., Masson, G. S., & Miles, F. A. (1997). Radial optic flow induces vergence eye movements with ultra-short latencies. *Nature*, 390, 512-515.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Coren, S., & Girgus, J. S. (1978). *Seeing is deceiving: The psychology of visual illusions*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Czerwinski, M., Tan, D. S., & Robertson, G. G. (2002). Women take a wider view. *Proceedings of CHI 2002*, 195-210.
- Dowling, J. E., & Boycott, B. B. (1966). Organization of the primate retina. *Proceedings of the Royal Society of London*, 16, 80-111.
- Frohlich, D., Helander, M., Landauer, T., & Prabhu, P. (1997). *Direct manipulation and other lessons, handbook of human-computer interaction*. North Holland: Elsevier Science Publishers.
- Gentilucci, M., Chieffi, S., Daprati, E., Saetti, M. C., & Toni, I. (1996). Visual illusion and action. *Neuropsychologia*, 34(6), 369-376.
- Goodale, M. A. (1994). One visual experience, many visual systems. *Attention and Performance XVI*.

- Goodale, M. A., & Humphrey, G. K. (1998). The objects of action and perception. *Cognition*, *67*, 181–207.
- Goodale, M. A., & Milner, A. (1992). Separate visual pathways for perception and action. *TINS*, *15*(1), 20–25.
- Grea, H., Pisella, L., Rossetti, Y., Prablanc, C., Desmurget, M., Tilikete, C., Grafton, S., & Vighetto, A. (2002). A lesion of the posterior parietal cortex disrupts on-line adjustments during aiming movements. *Neuropsychologia*.
- Haffenden, A., & Goodale, M. A. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neurosciences*, *10*, 122–136.
- Henson, D. B. (1993). *Visual fields*. Oxford: Oxford University Press.
- Hess, W. R., Burgi, S., & Bucher, V. (1946). Motor function of tectal and tegmental area. *Monatsschr. Psychiatr. Neurol.*, *112*, 1–52.
- Hinckley, K., Pausch, R., Goble, J. C., & Kassel, N. F. (1994). A survey of design issues in spatial input. *Proceedings of UIST '94*, 213–222.
- Hinckley, K., Tullio, J., Pausch, R., Proffitt, D., & Kassel, N. (1997). Usability analysis of 3d rotation techniques. *Proceedings of UIST '97*, 1–10.
- Holloway, R., & Lastra, A. (1993). *Virtual environments: A survey of the technology* (Technical Report No. TR93-033). Chapel Hill, NC 27599-3175: University of North Carolina at Chapel Hill.
- Ingle, D. (1973). Disinhibition of tectal neurons by pretectal lesions in the frog. *Science*, *180*, 422–424.
- Jackson, S. R. (2000). Perception, awareness, and action: Insights from blindsight. *Beyond Dissociation: Interaction Between Dissociated Implicit and Explicit Processing*, 73–98.
- Jacob, J. K., Sibert, L. E., McFarlane, D. C., & Mullen, M. P. (1994). Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction*, *1*(1), 3–26.
- Jacoby, R., Ferneau, M., & Humphries, J. (1994). Gestural interaction in a virtual environment. *Proceedings of Stereoscopic Display and Virtual Reality Systems: The Engineering Reality of Virtual Reality*, 355–364.

- Jakobson, L. S., Archibald, Y. M., Carey, D. P., & Goodale, M. A. (1991). A kinematic analysis of reaching and grasping movements in a patient recovering from optic ataxia. *Neuropsychologia*, *29*, 803–809.
- Jeannerod, M. (1986). The formation of finger grip during prehension: A cortically mediated visuomotor pattern. *Behavioural Brain Research*, *19*, 99–116.
- Jeannerod, M. (1999). A dichotomous visual brain? *Psyche*, *5*(25).
- Jeannerod, M., & Biguer, B. (1982). Visuomotor mechanisms in reaching within extrapersonal space. *Analysis of Visual Behaviour*.
- Kalawsky, R. S. (1993). Critical aspects of visually coupled systems. *Virtual Reality Systems*, 203–212.
- Lawton, W., Poston, T., & Serra, L. (1995). Time-lag reduction in a medical workbench. *Virtual Reality Applications*, 123–148.
- Lee, C., Ghyme, S., Park, C., & Wohn, K. (1998). The control of avatar motion using hand gesture. *Proceedings of VRST '98*, 59–66.
- Levine, M. D. (1985). *Vision in man and machine*. New York: McGraw-Hill.
- Liang, J. (1994). JDCAD: A highly interactive 3d modeling system. *Computers and Graphics*, *18*(4), 499–506.
- Liang, J., Shaw, C., & Green, M. (1991). On temporal-spatial realism in the virtual reality environment. *Proceedings of UIST '91*, 19–25.
- MacKenzie, I., & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. *Proceedings of INTERCHI '93*, 488–493.
- Mateeff, S., & Gourevich, A. (1983). Peripheral vision and perceived visual direction. *Biological Cybernetics*, *49*, 111–118.
- McGeorge, P. (1999). Consciousness, coordination, and two visual streams. *Psyche*, *5*(12).
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action, oxford psychology series* (Vol. 27). New York: Oxford University Press.
- Mine, M., Brooks, F., & Sequin, C. (1997). Moving objects in space: Exploiting proprioception in virtual environment interaction. *Proceedings of SIGGRAPH '97*, 19–26.

- Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., & Long, A. C. (2002). Interacting at a distance: Measuring the performance of laser pointers and other devices. *Proceedings of CHI 2002*, 33–40.
- Norman, D. (1988). *The psychology of everyday things*. Basic Books.
- Oh, J., & Stuerzlinger, W. (2002). Laser pointers as collaborative pointing devices. *Proceedings of Graphics Interface 2002*, 141–149.
- Olano, M., Cohen, J., Mine, M., & Bishop, G. (1995). Combatting rendering latency. *ACM Symposium on Interactive 3D Graphics*, 19–24.
- Olsen, D. R., & Nielsen, T. (2001). Laser pointer interaction. *Proceedings of CHI 2001*, 17–22.
- Osawa, N., Asai, K., & Sugimoto, Y. (2000). Immersive graph navigation using direct manipulation and gestures. *Proceedings of VRST 2000*, 147–152.
- Oviatt, S., & Cohen, P. (2002). Multimodal interfaces that process what comes naturally. *Communications of the ACM*, 43(3), 45–53.
- Paillard, J. (1991). Motor and representational framing of space. *Brain and Space, Chapter 10*, 163–182.
- Pelisson, D., Prablanc, C., Goodale, M. A., & Jeannerod, M. (1986). Visual control of reaching movements without vision of the limb. *Experimental Brain Research*, 62, 303–311.
- Perenin, M. T., & Rossetti, Y. (1996). Grasping in an hemianopic field. another instance of dissociation between perception and action. *Neuroreport*, 7(3), 793–797.
- Perenin, M. T., & Vighetto, A. (1988). Optic ataxia. *Brain*, 111(3), 643–674.
- Pisella, L., & Rossetti, Y. (2000). Interaction between conscious identification and non-conscious sensorimotor processing: Temporal constraints. *Beyond Dissociation: Interaction Between Dissociated Implicit and Explicit Processing*, 129–151.
- Poupyrev, I., Billinghamurst, M., Weghorst, S., & Ichikawa, T. (1996). Go-go interaction technique: Non-linear mapping for direct manipulation in vr. *Proceedings of UIST '96*, 79–80.

- Poupyrev, I., Weghorst, S., Billingham, M., & Ichikawa, T. (1998). Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Proceedings of Eurographics '98*, 1(3).
- Rauterberg, G. W. M. (2000). How to characterize a research line for user-system interaction. *IPO Annual Progress Report 35*, 66–86.
- Riggs, L. A. (1971). *Vision, experimental psychology*. New York: Holt, Rinehart, and Winston.
- Rizzolatti, G., Gentilucci, M., & Matelli, M. (1985). *Attention and performance xi*. Erlbaum.
- Roelofs, C. (1935). Optische localisation [optical localization]. *Archiv fur Augenheilkunde*, 109, 395–415.
- Rossetti, Y., & Pisella, L. (2002). Several ‘vision for action’ systems: A guide to dissociating and integratin dorsal and ventral functions. *Common Mechanisms in Perception and Action, Attention and Performance XIX*, 62–119.
- Schaufler, G., Mazuryk, T., & Schmalstieg, D. (1996). High fidelity for immersive displays. *Proceedings of CHI '96*.
- Schneider, G. E. (1969). Two visual systems: Brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions. *Science*, 163, 895–902.
- Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Shebilske, W. L. (1984). Context effects and efferent factors in perception and cognition. *Cognition and Motor Processes*, 99–119.
- Shneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. *Behaviour and Information Technology*, 1, 237–256.
- Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *IEEE Computer*, 16, 57–69.
- So, R. H., & Griffin, M. J. (1995a). Effects of lags on human operator transfer functions with head-coupled systems. *Aviat Space Environ Med*, 66, 550–556.
- So, R. H., & Griffin, M. J. (1995b). Head-coupled virtual environment with display lag. *Simulated and Virtual Realities: Elements of Perception*, 103–111.

- Song, C. G., Kwak, N. J., & Jeong, D. H. (2000). Developing an efficient technique of selection and manipulation in immersive v.e. *Proceedings of VRST 2000*, 142–146.
- Starner, S., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R., & Pentland, A. (1997). Augmented reality through wearable computing. *Presence*, 6(4), 386–398.
- Stoakley, R., Conway, M., & Pausch, R. (1995). Virtual reality on a wim: Interactive worlds in miniature. *Proceedings of CHI '95*, 265–272.
- Sutherland, I. E. (1965). The ultimate display. *Proceedings of the IFIP Congress*, 2, 506–508.
- Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, 31, 299–337.
- Ungerleider, L. G., & Mishkin, M. (1982). *Analysis of visual behaviour*. MIT Press.
- Ware, C., Arthur, K., & Booth, K. S. (1993). Fish tank virtual reality. *Proceedings of InterCHI '93: ACM Conference on Human Factors in Computing Systems*, 37–42.
- Ware, C., & Balakrishnan, R. (1994). Reaching for objects in vr displays: Lag and frame rate. *ACM Transactions on Computer-Human Interaction*, 1(4), 331–356.
- Watson, B., Walker, N., Ribarsky, W., & Spaulding, V. (1998). Effects of variation in system responsiveness on user performance in virtual environments. *Human Factors, Special Section on Virtual Environments*, 40(3), 403–414.
- Weiskrantz, L. (1986). *Blindsight: A case study and implications*. New York: Oxford University Press.
- Westwood, D. A., Heath, M., & Roy, E. A. (2000). The effect of a pictorial illusion on closed-loop and open-loop prehension. *Experimental Brain Research*, 134, 456–463.
- Wexelblat, A. (1995). An approach to natural gesture in virtual environments. *ACM Transactions on Computer-Human Interaction*, 2(3), 179–200.

Appendix A

Audio Recording Script for Subject Instructions

In order to record the necessary audio instructions for use during the two visual systems experiment in virtual reality, the following text script was prepared. Each individual portion of the script was stored as an independent WAV file, and was played back via the experiment software application to participating subjects at appropriate moments during the experiment sessions. Note that this script is not presented in an order that is indicative of the order in which the individual portions were presented during the experiment.

A.1 General Notes and Instructions to Subjects

You are about to participate in an experiment that has been designed to examine some relationships between visual perception and motor action in large scale virtual reality systems. Please listen to the following notes carefully. If you have any questions, please feel free to ask me before we begin.

- This experiment is expected to be approximately sixty to ninety minutes in duration. There will be several opportunities for you to take breaks during

the course of the experiment.

- You will be given specific instructions prior to each part of the experiment. Please make sure that you listen to these instructions carefully.
- I will be present at all times during the experiment. Please feel free to let me know if you have any questions or concerns about this experiment.
- Remember that you are free to withdraw from this experiment at any time. Please be sure to let me know if you do not wish to continue to participate.

Whenever you feel ready, let me know and we'll begin the experiment. Thank you for participating!

A.2 Cognitive Report Condition Instructions

In this part of the experiment, you will be presented with a series of trials that will require you to make verbal responses. For each trial in this part of the experiment:

- A circular target will appear, then disappear after a short period of time. Some of these targets will be surrounded by rectangular frames. In addition, you will be given verbal feedback for some of these trials, while for other trials you will be given no feedback.
- For each circular target that is presented, it will be your task to verbally tell me where you think each circular target was, in relation to straight ahead. Your response will be one of five possible choices: “Far Left,” “Left,” “Centre,” “Right,” and “Far Right.”
- You will hear me say “Respond!” when I would like you to give me an answer. Don't respond before I give you this indication to do so.
- Keep your head completely still, and keep your arms down during this part of the experiment.

If you have any questions, please feel free to ask me before we begin. When you are ready to begin this part of the experiment, let me know.

A.3 Open Loop Pointing Condition Instructions

In this part of the experiment, you will be presented with a series of trials that will require you to make physical pointing gestures through the use of the stylus that has been given to you. For each trial in this part of the experiment:

- A circular target will appear, then disappear after a short period of time. Some of these targets will be surrounded by rectangular frames. In addition, you will be given verbal feedback for some of these trials, while for other trials you will be given no feedback. Some trials will also give you an on-screen crosshair that follows your pointing movements, while other trials will give you no such guide.
- For each circular target that is presented, it will be your task to acquire the target's presented position by pointing at where the circular target was presented. To point with the stylus, point at the target's position, and when you are satisfied with where you are pointing, hold the stylus in place until you hear me say "Trial Complete." When you hear me say this, return the stylus to a resting position on your lap, then press the button on the stylus when you are ready to advance to the next trial.
- You will hear me say "Point now!" when I would like you to acquire the target's last position. Don't point before I give you this indication to do so.
- Keep your arms underneath the wooden frame at all times. Keep your head completely still during this part of the experiment.

If you have any questions, please feel free to ask me before we begin. When you are ready to begin this part of the experiment, let me know.

A.4 Closed Loop Pointing (Lag and No Lag) Condition

Instructions

In this part of the experiment, you will be presented with a series of trials that will require you to make physical pointing gestures through the use of the stylus that has been given to you. For each trial in this part of the experiment:

- A circular target will appear, then disappear after a short period of time. Some of these targets will be surrounded by rectangular frames. In addition, you will be given verbal feedback for some of these trials, while for other trials you will be given no feedback. In this part of the experiment, you will see an on-screen crosshair that tracks your pointing movements.
- For each circular target that is presented, it will be your task to acquire the target's presented position by pointing at where the circular target was presented. To point with the stylus, point at the target's position, and when you are satisfied with where you are pointing, hold the stylus in place until you hear me say "Trial Complete." When you hear me say this, return the stylus to a resting position on your lap, then press the button on the stylus when you are ready to advance to the next trial.
- You will hear me say "Point now!" when I would like you to acquire the target's last position. Don't point before I give you this indication to do so.
- Keep your arms underneath the wooden frame at all times. Keep your head completely still during this part of the experiment.

If you have any questions, please feel free to ask me before we begin. When you are ready to begin this part of the experiment, let me know.

A.5 Cognitive Report Target Presentation Notes

What you are now seeing are the five possible locations at which you will see targets in this part of the experiment. Please take a moment to familiarize yourself with these locations. Pay particular attention to the labels beneath the targets. These are the labels that you will call out to me when you are asked to respond.

For example, if a target appears to be presented directly straight ahead, you will respond with the answer “Centre.” If a target appears to be presented in the furthest left position, you will respond with the answer “Far Left” and so on. When you feel comfortable enough with your knowledge of the target positions, please let me know, and we will continue with the experiment.

A.6 Shared Pointing Target Presentation Notes

What you are now seeing are the five possible locations at which you will see targets in this part of the experiment. Please take a moment to familiarize yourself with these locations. When you feel comfortable enough with your knowledge of the target positions, please let me know, and we will continue with the experiment.

A.7 Cognitive Report Trial Progression Notes

- Let’s begin this experiment with just targets on the screen. Following each trial, you will be given verbal feedback that indicates the correctness of your trial responses. Get ready!
- Good! Now, let’s add rectangular frames to the targets that are being presented. You will continue to be provided with verbal feedback that indicates the correctness of your trial responses. Get ready!
- Great! Let’s keep going with these target and frame presentations, except that you will no longer be given any verbal feedback after each trial. Get ready!

A.8 Open Loop Pointing Trial Progression Notes

- Great! Let's keep going with these target and frame presentations, except that you will no longer be presented with an on-screen crosshair. You will still be provided with verbal feedback that indicates the correctness of your trial responses. Get ready!
- Good! Let's keep going with these target and frame presentations, except that you will also no longer be given any verbal feedback after each trial. Get ready!

A.9 Shared Pointing Trial Progression Notes

- Let's begin with just targets on the screen. Following each trial, you will be given verbal feedback that indicates the correctness of your trial responses. An on-screen crosshair that follows your pointing motions will also be present. Get ready!
- Good! Now, let's add rectangular frames to the targets that are being presented. You will continue to be provided with verbal feedback that indicates the correctness of your trial responses, and you will continue to be presented with an on-screen crosshair. Get ready!
- Good! Let's keep going with these target and frame presentations, except that you will no longer be given any verbal feedback after each trial. You will still continue to be presented with an on-screen crosshair that tracks your pointing motions. Get ready!

A.10 Condition Concluding Notes

Please take a moment to relax. If you have any questions, comments, or concerns at this time, now would be a good time to discuss them. Whenever you feel that

you are ready to continue with the next part of the experiment, please let me know.

A.11 Experiment Session Concluding Notes

This experiment is now complete. I will help you remove and detach any equipment, then provide you with a session debriefing that will disclose the purpose and motivation for this experiment.

I would like to remind you that all of the results and data that have been collected during this session will be kept strictly anonymous, and that your identity will not be divulged in any way. If there is anything that you would like to know, please feel free to ask me.

Thank you again for your participation!

A.12 Audio Indications and Feedback

A.12.1 Verbal Indications to Respond

- Respond!
- Point now!
- Trial complete!

A.12.2 Practice Trial Verbal Feedback

- That's correct!
- That's incorrect. The target was actually "Far Left."
- That's incorrect. The target was actually "Left."
- That's incorrect. The target was actually "Centre."
- That's incorrect. The target was actually "Right."

- That's incorrect. The target was actually "Far Right."
- This red circle indicates where the target was, and this yellow circle indicates where you pointed.