

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Perception-Action Loop in the Experience of Virtual Environments

Permalink

<https://escholarship.org/uc/item/96v4q8g9>

Author

Kim, Seung Wook

Publication Date

2009

Peer reviewed|Thesis/dissertation

Perception-Action Loop in the Experience of Virtual Environments

by

Seung Wook Kim

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Architecture

and the Designated Emphasis

in

New Media

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yehuda E. Kalay, Chair

Professor Galen Cranz

Professor John F. Canny

Fall 2009

Abstract

Perception-Action Loop in the Experience of Virtual Environments

by

Seung Wook Kim

Doctor of Philosophy in Architecture

and the Designated Emphasis in New Media

University of California, Berkeley

Professor Yehuda Kalay, Chair

The goal of this work is to develop an approach to the design of natural and immersive interaction methods for three-dimensional virtual environments. My thesis is that habitation and presence in any environment are based on a continuous process of perception and action between a person and his/her surroundings. The current practice of virtual environments, however, disconnects this intrinsic loop, separating perception and action into two different ‘worlds’—a physical one (for perception) and a virtual one (for action). This research is aimed at bridging the gap between those two worlds.

Being drawn from perceptual philosophy and psychology, the theoretical study in this dissertation identifies three embodiments of natural perception-action loop: direct perceptual acts, proprioceptive locomotion, and motor intentionality. These concepts form the basis for the interaction methods proposed in this work, and I demonstrate these methods by implementing pertinent prototype systems. First, I suggest a view-dependent, non-planar display space that supports natural perceptual actions, thereby enhancing our field of view as well as depth perception. Second, I propose Interactive Chair, which provides an intuitive locomotion control based on our sense of self-body movements. Third, I argue that pointing-based interaction techniques inhibit our motor-intentional behaviors, therefore demonstrate a line-of-sight, direct object manipulation method. I finally conduct a series of user tests to evaluate the proposed methods and systems, and confirm the contribution of suggested interaction design approaches to the natural experience of virtual environments.

ACKNOWLEDGEMENT

Since I arrived at Berkeley in the summer of 2002, Professor Yehuda E. Kalay has been a great mentor, continually and convincingly conveying a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. Without his guidance and persistent help this dissertation would not have been possible. I would like to express the deepest appreciation to him.

Special thanks to my committee members, Professor Galen Cranz and Professor John F. Canny for their advice and encouragement. Dr. Cranz was also my qualifying examination chair, and Dr. Canny was the advisor for my master's study in computer science as well.

I also would like to thank my qualifying examination committee members, Professor Alva Noë in Philosophy and Professor Anthony Burke in Architecture for their supports.

Digital Design Research Group has been my home base at UC Berkeley. I wish to thank all my colleagues, including Wei Yan, Lily Chen, Jaewook Lee, Yongwook Jeong, Cecilia Kim, Gokce Kinayoglu, Emelie Cheng, Selina Lam, Bauni Hamid, Ahmed El Antably, Yael Perez, Hsin-Hsien Chiu, Kartikeya Date, and Seung Wan Hong.

A thank you to Professor Kimiko Ryokai in Information School, who introduced me to, not only the study of tangible user interface, but also my current workplace, Samsung R&D Center at Silicon Valley.

In addition, I would like to thank to my colleagues at work, especially my manager Victoria Coleman who offered me a great job position, and my team head Stefan Marti who gave me a great deal of inspiration in regard to 3D UI research.

My father, Dong Whan Kim, and my mother Eui Sun Choi, have been always supporting me to study abroad from their heart. My big sister Siyeon Kim and her husband Sungsook Kang, and my other sister Doyon Kim and her husband Hakpyo Hong, have been also great helpers for me.

Finally, I would like to thank my wife, Jee Hyun Yang, who is also a mother of two lovely daughters, Suhyeon Kim and Soohyang Kim, for her dedication to my doctoral study over many years. Her encouragement cheered me through the entire process.

CONTENTS

Chapter 1 Introduction.....	1
1.1 The Problem: Gap between Perception and Action	1
1.2 The Structure of This Dissertation	2
Chapter 2 Relationship between Perception and Action.....	3
2.1 Immersive Presence in Virtual Environments	3
2.1.1 Cartesian dualism and virtual presence	3
2.1.2 Heideggerian existentialism.....	4
2.2 Perception and Body	5
2.2.1 Indirect vs. direct realism.....	5
2.2.2 Phenomenology of perception	6
2.3 Relationship between Perception and Action	8
Chapter 3 Direct Perceptual Acts in Virtual Space.....	9
3.1 Perceptual Acts as Sensorimotor Skills	9
3.2 System Supporting Direct Perceptual Acts.....	10
3.3 Design and Implementation of Interactive Non-Planar Display System	13
3.3.1 Non-planar display space design	13
3.3.2 Motion parallax and view-dependent rendering (VDR).....	15
3.3.2.1 Face tracking with OpenCV.....	15
3.3.2.2 Asymmetric view frustum cluster	16
3.3.3 Horizontal extension of non-planar display space.....	19
3.4 Discussion: VDR and Stereoscopic Displays	19
Chapter 4 Interactive Chair: Proprioceptive Locomotion Control	21
4.1 Proprioception and Locomotion	21
4.2 Human Body and Spatial Concepts	22
4.3 Technologies for Locomotion Control in 3D Space	23
4.4 Design and Implementation of Interactive Chair.....	24
4.4.1 Spatial correlation between sitting postures and 3D locomotion	24
4.4.1.1 Leaning behaviors and locomotion	24
4.4.1.2 Swiveling and orientation	25
4.5 Sensing Transitional Sitting Postures	26
4.5.1 Sensor arrangement and signal processing	27
4.5.2 Data mapping in the 3D engine	28
4.6 Swiveling Motions and Optical Flow	28
4.7 Discussion	29

Chapter 5 Motor-Intentional Manipulation of Virtual Objects	31
5.1 Grasping, Pointing and Motor Intentionality	31
5.2 Interaction Techniques for Virtual Object Manipulation.....	32
5.2.1 Pointing as an object selection technique	32
5.2.2 Virtual hand techniques	34
5.3 Direct and Motor-Intentional Manipulation of Virtual Objects.....	35
5.3.1 Finger Tracking and Hand Pose Detection	36
5.3.2 In-line Coordination between Face, Hand and Virtual Object	37
5.4 Discussion	38
Chapter 6 User Test and Evaluation	39
6.1 User Test Overview	39
6.2 Experiment 1: View-Dependent 3D Content in Non-Planar Display Space	41
6.2.1 Test Procedure	42
6.2.2 Performance Evaluation.....	42
6.2.3 Questionnaire Evaluation.....	43
6.3 Experiment 2: Proprioceptive Locomotion Control with Interactive Chair	44
6.3.1 Test Procedure	44
6.3.2 Performance Evaluation.....	46
6.3.3 Questionnaire Evaluation.....	46
6.4 Experiment 3: Motor-Intentional, In-line Manipulation of Virtual Objects	47
6.4.1 Test Procedure	47
6.4.2 Performance Evaluation.....	48
6.4.3 Questionnaire Evaluation.....	49
6.5 Discussion	49
Chapter 7 Conclusion	51
7.1 Conclusion	51
7.2 Future research directions	51
References	53

List of Figures

3.1 Head-mounted displays (HMDs)	11
3.2 CAVE (CAVE Automatic Virtual Environment) systems	11
3.3 Systems supporting head movements	12
3.4 FOVs in non-planar and planar display spaces.....	13
3.5 A prototype vertical non-planar display with three monitors	14
3.6 View-dependent rendering and motion parallax.....	15
3.7 Face tracking test with OpenCV library	16
3.8 Comparison between single (top) and multiple (bottom) view frustums.....	17
3.9 Symmetric vs. asymmetric view frustums	17
3.10 Calculation of asymmetric view frustum in a non-planar display space	18
3.11 Horizontal extension of display space and view frustums.....	19
4.1 CyberCarpet (left) and VirtuSphere (right).....	23
4.2 Leaning forward/backward with Interactive Chair	25
4.3 Swiveling with Interactive Chair	25
4.4 SensingChair and sitting pattern image	26
4.5 Prototype Interactive Chair and sensor systems	27
4.6 Portable sensor box configuration	28
4.7 Feature extraction and optical flow vector.....	29
5.1 Virtual ray-casting technique.....	33
5.2 Aperture selection technique.....	34
5.3 Oblong Industries' G-Speak technology	35
5.4 In-Line mediation of object manipulation	35
5.5 Finger tracking and hand poses.....	36
5.6 In-line coordination between face, hand and virtual object.....	37
6.1 Experimental systems (from left: for experiment 1, 2 and 3).....	41
6.2 Experiment 1 performance comparison	43
6.3 Experiment 1 questionnaire (T-test)	44
6.4 Route for virtual traveling.....	45
6.5 Experiment 2 performance comparison	46
6.6 Experiment 2 questionnaire (T-test)	47
6.7 Experiment 3 performance comparison	48
6.8 Experiment 3 questionnaire (T-test)	49

List of Tables

6.1 Post-questionnaire items in three categories.....	40
6.2 Mouse mapping in the controlled system	41

CHAPTER 1 INTRODUCTION

John R. Searle, describing what makes our experience of the world successful, said: “One of the chief functions of the mind, both in our day-to-day living and over the long evolutionary haul, is to relate us to the rest of the world, especially by way of perception and action”.¹ He further elaborated that we take in information about the world by perception, we then coordinate this information both consciously and unconsciously, and make decisions or otherwise form intentions, which result in actions by way of which we cope with the world.

The same principle – that perception and action are the keys to our presence in the physical world – is likely to hold true in virtual environments as well:

*Presence is tied to one’s successfully supported action in the environment, this environment being either virtual or real. The coupling between perception and action is crucial for determining the extent to which actions are successfully supported... But what specifically is meant by successfully supported action? When actions are made in an environment, the environment reacts, in some fashion, to the action made. When the environmental response is perceived as lawful, that is, commensurate with the response that would be made by the real world environment in which our perceptual systems have evolved, then the action is said to successfully support our expectations.*²

Accordingly, the success of virtual environments depends on how well this continuous loop between perception and action can be established across the line of demarcation between physical and virtual territories. If it is poorly designed, the virtual world may feel like a distant universe with extravagant geometries: a surreal space that attracts people once and is hardly revisited afterward. If the loop is well designed, on the other hand, the virtual world will be able to enrich interactions and relationships between its occupants, so that they can truly inhabit this digitally-mediated environment.

This dissertation is about the design of interaction methods for virtual environments: it is about the design of 3D user interfaces that allow people to inhabit virtual environments by naturally perceiving and acting. The premise in this research is that habitation and presence (in any environment) are based on a continuous process of perception and action between a person and his/her surroundings. It will question what happens when perception and action are separated into two different ‘worlds’—a physical one (for perception) and a virtual one (for action)?

1.1 The Problem: Gap between Perception and Action

Virtual Reality, made possible by three-dimensional computer graphics, is now merging with internet technology to allow social inhabitation of computer-generated virtual worlds. In its most successful form, multi-user video games have begun to provide possibilities for socializing and communal action in virtual environments for entertainment purposes. Extending that capability to include other types of activities is an ongoing area of research worldwide.

Scholars, nevertheless, have been skeptical of accepting this illusory but pervasive phenomenon in simulated, virtual worlds as a legitimate extension of social interaction in physical environments. The absence of physical embodiment is often cited as the most important

¹ Searle, J. R. (2004). *Mind: A Brief Introduction*. Oxford University Press, 259.

² Zahorik, P. and Jenison, R. L. (1998). Presence as Being-in-the-World. In *Presence: Teleoperators and Virtual Environments*, (7) 1: 78-89.

reason, raising questions about the nature of ‘presence’ in virtual environments. Normally, our physically-embodied actions are directed by immersion in the environment itself and thus bypassing the cognitive faculties, which are, however, called for by the kind of disembodied actions in case of interacting with virtual worlds. Every aspect of the action must be controlled by the human user, through visual perception of the virtual environment, and directing the avatar’s actions from ‘outside’ that environment. Such separation caused by immersion in a physical world and action in a virtual world creates an incompatibility between the perception of the environment and reaction to it—a disconnection in the perception-action loop.

By bridging this gap and providing seamless interface between physical and virtual worlds, this dissertation is aimed to make our presence in and inhabitation of virtual environments more natural and immersive. From an architectural point of view, this study pertains to the possibility of engendering a ‘sense of place’ in users who interact and socialize with others in virtual environments.³ From a human-computer interaction (HCI) point of view, this study pertains to the possibility of providing a smooth and natural interface between the user and his/her avatar, as well as the virtual environment in which that avatar exists. This dissertation will include both the theoretical investigation and the development of prototypical user interface systems.

1.2 The Structure of This Dissertation

In the following chapter 2, I will begin with a series of philosophical inquiries about meaning of presence, perception and action, and will identify that the coupling between perception and action is crucial to the veridical sense of presence. I will then suggest three embodiments of specific perception-action relationships – 1) direct perceptual acts, 2) proprioceptive locomotion and 3) motor intentionality, each of which will be investigated in the next three chapters.

First, in the chapter 3, I emphasize the importance of skillful, bodily movements in perceiving the world, and propose an interaction method based on a non-planar display space and a view-dependent rendering technique. Second, in the chapter 4, I pay attention to the relation between the sense of self-body movement and locomotion, and demonstrate Interactive Chair as an efficient approach to utilize our torso movements for locomotion control in the virtual space. Third, in the chapter 5, I point out that the current practices of virtual object manipulation mostly rely on the user’s cognitive process, while our interaction with physical objects is rather motor-intentional. I then suggest a method for direct manipulation of virtual objects, which is based on the line-of-sight coordination between a hand and a virtual object.

Finally, in the chapter 6, I perform a series of user tests, and evaluate their results to validate suggested prototype systems. In the chapter 7, I conclude this dissertation work.

³ Kalay, Y. E. (2004). *Architecture’s New Media*. MIT Press.

CHAPTER 2 RELATIONSHIP BETWEEN PERCEPTION AND ACTION

2.1 Immersive Presence in Virtual Environments

Three-dimensional virtual worlds have grown beyond the vision of *Metaverse*,¹ to open up opportunities for immersive presence in digital environments. While immersion is understood as a description of overall fidelity in relation to physical reality provided by a medium,² presence, in this context, is typically defined as the perceptual illusion of non-mediation.³ Feeling of immersive presence is a driving force that attracts people into such a virtual place, providing socio-culturally rich experiences.⁴ In recent years, both academic and industrial societies have been deeply involved in the effort to engender an immersive sense of presence in virtual environments. However, the advance in digital media technologies has ruled out the physicality of our existence in experiencing such mediated worlds, only taking our mind into the virtual territory. This has raised philosophical questions and insights, which has entailed re-conceptualizing the nature of presence.

2.1.1 Cartesian dualism and virtual presence

Since the early 1990's, a growing number of scholars have turned their attention to presence in the virtual world. In their efforts to create an interdisciplinary framework, researchers have struggled with the philosophical grounding of presence. Particularly, they often consider the French philosopher René Descartes as the modern thinker who established the basis for research on presence. Descartes conceived a human being as a thinking thing, and postulated a dualism between *res extensa* (objects located outside the mind) and *res cogitans* (objects located within the mind).⁵ Modern studies on presence subscribe, in one way or another, to this Cartesian view – that there are clearly separable mental and physical domains, as it pervades today's physical science, engineering and cognitive science.⁶ Zahorik and Jenison point out that, the understanding of presence has traditionally relied on the rationalistic orientation of such dualism.⁷ Rationalism holds that our knowledge is obtained on the basis of reason (or rationality), and that there exists a rational process called 'representation' which transduces the knowledge of external world into the subjective experience. This promotes the understanding of presence as the representational function to bridge the gap between those mental and physical domains. However Descartes did not deny the possibility of an unperceived mediation that makes our presence fictional. Going through a skeptical reasoning, he speculated on a fundamental tension between

¹ The Metaverse is a fictional virtual world, described in Neal Stephenson's 1992 science fiction novel *Snow Crash*, where humans, as avatars, interact with each other and software agents, in a three-dimensional space that uses the metaphor of the real world; Stephenson, N. (1992). *Snow Crash*. Bantam Books.

² Slater, M. (2002). Course Notes on Understanding Virtual Environments: Immersion, Presence and Performance. In *ACM SIGGRAPH*, San Antonio, TX.

³ Lombard, M. and Ditton, T. (1997). A the heart of it all: The concept of presence. In *Journal of Computer-Mediated Communication*, 3(2). WWW URL: <http://jcmc.indiana.edu/vol3/issue2/lombard.html>

⁴ Kalay, Y. E. (2004). *Architecture's New Media*. MIT Press.

⁵ Descartes, R. (1984 [1641]). Meditations on First Philosophy. In Cottingham, J., Stoothoff, R. and Murdoch, D. (Trans.). *The Philosophical Writings of René Descartes*, Cambridge University Press, (2): 1-62.

⁶ Sheridan, T. B. (1999). Descartes, Heidegger, Gibson, and God: Toward an Eclectic Ontology of Presence. In *Presence: Teleoperators and Virtual Environments*, (8) 5: 551-559.

⁷ Zahorik, P. and Jenison, R. L. (1998). Presence as Being-in-the-World. In *Presence: Teleoperators and Virtual Environments*, (7) 1: 78-89.

the actual experience of something (e.g., Descartes being in his room) and the unreliable perception of experience (e.g., Descartes dreaming or being misled by a malicious demon).

In this context, virtual presence can be understood as Cartesian failure to recognize the technologically-mediated nature of our experiences.⁸ With this Cartesian approach, the notion of presence is reduced to a type of perception, for example, visual perception, since the stress is on the possibility of epistemic failure to perceive the difference between what is and what is not real. When the powerful mediation technologies replace Cartesian dreams and demons, the condition of presence comes down to the matter of failing to perceive those technologies that make one dream. Modern endeavors to promote virtual presence often build on this conceptual model of presence.

2.1.2 Heideggerian existentialism

By strictly separating mental and physical domains, Descartes claims that only the mental contents can be immediate and self-evident, thus the very act of probing these mental contents (thinking) guarantees our existence. Martin Heidegger, a German philosopher, contended against this conception, as he was principally concerned with the question of what it means *to be* (to exist). Heidegger found Descartes' account unsatisfactory in describing the meaning of human existence, because 'meaning' is contingent upon interpretation, and interpretation is always biased by the interpreter's beliefs, language, practices, and so on. According to him, we are "thrown" into situations in which we must continually interpret, and we all exist in the world in this type of thrown state.⁹ We are all continually acting and therefore not able to represent the situation at hand in a detached and analytic manner.¹⁰ In this mode of "being," we do not have stable mental representations of our environment, but do have complete involvement in a dynamic interaction, and only by stepping back and disconnecting from that involvement can a person perceive the elements of the situation.¹¹ Following this Heideggerian existentialism, Dreyfus argues that Cartesian representations are not impossible but are not primary modes of presence either:

*We should try and impress on ourselves what a huge amount of our lives – dressing, working, getting around, talking, eating, etc. – is spent in this state (the "thrown," non-analytic state), and what a small part is spent in the deliberate, effortful, subject/object mode, which is, of course, the mode we tend to notice, and which has therefore been studied in detail by philosophers.*¹²

Dourish explains this with an example of a computer mouse:

Much of the times, I act through the mouse; the mouse is an extension of my hand as I select objects, operate menus, and so forth. The mouse is, in Heidegger's terms, ready-to-hand. Sometimes, however, such as when I reach the edge of the mousepad and cannot move the mouse further, my orientation toward the mouse changes. ... The mouse becomes the object of my attention as I pick it up and move it back to the center of the

⁸ Floridi, L. (2005). The Philosophy of Presence: From Epistemic Failure to Successful Observation. In *Presence: Teleoperators and Virtual Environments*, (14) 6: 656-667.

⁹ Heidegger, M. (1962). *Being and time* (Macquarrie, J. and Robinson, E. Trans.). Harper Collins, San Francisco.

¹⁰ Zahorik and Jenison, Op.cit.

¹¹ Sheridan, Op.cit.

¹² Dreyfus, H. L. (1991). *Being-in-the-world: A commentary on Heidegger's being and time, division I*. MIT Press, Cambridge, MA.

*mousepad. When I act on the mouse in this way, being mindful of it as an object of my activity, the mouse is present-at-hand.*¹³

Based on this distinction between “ready-to-hand” and “present-at-hand”, Zahorik and Jenison construe that being is defined in terms of action within a worldly context, and that existence is action and action is existence. Finally, they underline that “*Presence is tantamount to successfully supported action in the environment*”.¹⁴ In the case that we act through (mediation) technology which has become ready-to-hand, it disappears from our immediate concerns and fades into the background. In Dourish’s account, this unspoken background against which our actions are played out is at the heart of Heidegger’s view of being-in-the-world.¹⁵ In this ‘existentialism,’ after all, the notion of presence is not merely reduced to perception, but is rather reconstructed as a kind of interplay between an environment, a perceiver, and her actions in that environment.

2.2 Perception and Body

Whether Cartesian or Heideggerian thinking, our perceptual experience seems to be at the center of being-in-the-world. However, some philosophers maintain that we cannot experience the presence of a thing or event unless we perceive it in relation to the position of our own bodies, raising questions about the nature of presence in an environment, such as, ‘how do we perceive a thing or event?’ ‘what is the role of our bodies in such a perceptual experience?’ and so on. Even though answers to these questions will always remain debatable we still need to investigate how this discussion has been evolved over history, given their potential influences on the design of virtual environments.

2.2.1 Indirect vs. direct realism

We may believe that what we perceive in the world are things in themselves, rather than patterns that overlay the forms of those things. However, Descartes was convinced that this naïve realism is false. By studying sense organs, he figured out how the eye responds to light and passes the information on to the brain through fibers of the optic nerve, and concluded that our access to the world is indirect, thus things are never directly present to us.¹⁶ Searle found that Descartes’ contemporary thinkers such as Locke, Hume and Kant also believed that this ‘indirect’ realism is true, describing it in their own terms, “idea”, “impression” and “representations”, respectively.¹⁷ Moreover, philosophers in the past century put this point by saying “we do not perceive material objects, we perceive only sense data.” The British philosopher George Moore, who introduces this sense-data theory, and his followers Russell and Price, consistently came to the conclusion that the ‘sense data’ are not objects themselves, but something else which is dependent on the mind.¹⁸ They denied that material objects were the things directly given to the mind because of the notorious ‘argument from illusion’.¹⁹ Illusion, or hallucination, shows that one never directly

¹³ Dourish, P. (2004). *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, MA, 109.

¹⁴ Zahorik and Jenison, Op.cit.

¹⁵ Dourish, Op.cit.

¹⁶ Descartes, R. (1958). Dioptric. In *Descartes: Philosophical Writings*. (Smith, N. K. Ed. and Trans.). New York.

¹⁷ Searle, J. R. (2004). *Mind: A Brief Introduction*. Oxford University Press, 260.

¹⁸ *Stanford Encyclopedia of Philosophy*. WWW URL: <http://plato.stanford.edu/entries/moore/>

¹⁹ Crane, T. (2001). *Elements of mind: an introduction to the philosophy of mind*. Oxford University Press, 132.

or immediately perceives a material object, since one's experience could remain the same even if there were no such object perceived. This argument has appeared in many forms throughout the history of philosophy.

On the contrary, Thomas Reid, a Scottish philosopher in the eighteenth century, argued that our minds connect to the world directly, rather than through some sort of medium (such as 'idea' by Locke). He suggests that perception is a species of conception, and to perceive an object is to be aware of it in a particular way to be convinced that the object exists and is as you conceive it to be. Sensations, such as touch or vision, lead us to the conception of their causes, and therefore help us to place our minds in direct contact with the world.²⁰ This 'direct realism,' however, does not entirely advocate the more naïve realism. While the naïve view considers the world is exactly as we perceive it to be, the direct realism simply claims that what we directly perceive is in the external world rather than in the mind. Accordingly, direct realism does not completely deny hallucinatory cases of perceptual experiences, as opposed to veridical ones. Therefore, the argument from illusion can still be taken as an argument against direct realism.

In the area of perceptual psychology, James J. Gibson was principally concerned with direct visual perception. He agreed that the function of vision is to keep the perceiver in touch with the environment and to guide action, not to produce inner experiences and representations. In his theory, perception is a direct process of picking up information from an informationally rich environments. Therefore, perception is not an occurrence that takes place in the brain of the perceiver, but rather is an act of the whole animal, the act of perceptually guided exploration of the environment²¹. The environment is not just the object of perception, but also the source of perceptual information, which can be considered as properties that uniquely specify that environment, rather than merely proximal stimulation. Gibson termed this action-supportive information an *affordance*.²² In contrast to the indirect realism in which perception is veridical if it matches the states of affairs in the real objective world, Gibson proposed that perception is veridical to the extent that it supports successful action in the environment. He, thus, placed visual perception within a frame of being and acting, and in doing so laid the foundations for what he and others came to call "ecological psychology."²³ This view immediately invalidates illusionary or hallucinatory experiences as they are incidents isolated from the reciprocal process between our actions and the world.

2.2.2 Phenomenology of perception

Gibson's direct visual perception is in line with the 'thrown state of being' introduced by Heidegger, in that both of them relate us (as perceivers) to our given situation (environment) by means of acting. The three-way relationship between environment, perception and action, mentioned in the beginning of this section, now tells us that we gain knowledge of (i.e., perceive) an environment, through that environment's responses to the actions we made. A French phenomenologist Maurice Merleau-Ponty reconfirms this action-environment feedback loop in our being-in-the-world, further stressing that it is our own body which keeps running that loop:

²⁰ *Stanford Encyclopedia of Philosophy*. WWW URL: <http://plato.stanford.edu/entries/reid/>

²¹ Noë, A. (2002). Is the Visual World a Grand Illusion? In *Journal of Consciousness Studies*, 9 (5-6): 1-12.

²² Gibson, J. J. (1979). *The ecological approach to visual perception*, Hillsdale, NJ.

²³ Dourish, Op.cit., 118.

*My body is geared into the world when my perception presents me with a spectacle as varied and as clearly articulated as possible, and when my motor intentions, as they unfold, receive the responses they expect from the world. This maximum sharpness of perception and action points clearly to a perceptual ground, a basis of my life, a general setting in which my body can co-exist with the world.*²⁴

In his phenomenological account, the unity of the world is correlated with our body's unity. When we see a cube with six equal sides, for instance, we cannot grasp the unity of the cube without the mediation of bodily experience. In this case, the appearance of the cube is presented in different perspectives, but we do not construct the idea of the flat projection which accounts for these perspectives. Merleau-Ponty conceives that, because the new appearance has already compounded itself with the live-through movement and presented itself as an appearance of a cube, there is no need to take an objective view of our own movement in order to reconstitute the true form of the cube behind its appearance; the perception of the world and our bodily movements are rather inseparable:

*The thing, and the world, are given to me along with the parts of my body, not by any "natural geometry," but in a living connection comparable, or rather identical, with that existing between the parts of my body itself.*²⁵

To Merleau-Ponty, however, being able to act on and get feedback from the world is not sufficient to get a sense of direct contact with reality. While the perceptual environment sought in the ecological approach is informationally rich and affordable, Dreyfus construes that the perceptual world in Merleau-Ponty's phenomenology is rather uncertain and unstable, and for this reason we constantly move to overcome such uncertainty and instability so as to make our perceptual experience more stable. In doing so, our body keeps soliciting a readiness to get both an optimal "grip" on things as well as a sense of overall context in their surroundings:²⁶

*There is thus a constantly enriched interaction between the details and the overall significance of the situation. Merleau-Ponty calls this kind of feedback between one's actions and the perceptual world, the intentional arc. And he points out that it functions only if the perceiver is using his body as an 'I can', that is, in this case, if he controls where he looks.*²⁷

What the possession of physical body affords to us, as embodied beings, is then the abilities of bodily skills, and these skills constantly keep up with our inherent perceptual demands against the environment. When such skillful actions keep fulfilling our perception, and in turn that perception keeps leading to our next action, the state of our being-in-the-world becomes stable. Therefore, the feedback between actions and the environment is not autonomous, but radically intended by our perceptual needs. Being enabled by our body, a more intrinsic feedback loop is running between our perception and action. In this sense, *the theory of the body is already a theory of perception.*²⁸

²⁴ Merleau-Ponty, M. (2003 [1945]). *Phenomenology of Perception*. Routledge Classics, 292.

²⁵ Ibid.

²⁶ Dreyfus, H. L. (2001). *Think in Action: On the Internet*. Routledge, 55.

²⁷ Ibid., 66.

²⁸ Merleau-Ponty, Op.cit., 235.

2.3 Relationship between Perception and Action

We have seen from philosophical inquiries that the coupling between perception and action is crucial to the veridical sense of presence. In virtual environments, however, our perception and action are separated into two different ‘worlds’—a physical one (for perception) and a virtual one (for action): what we (intend to) do inside the screen can be observed only outside the screen. Thus, it may sound contradictory to bridge the gap between those two. Phenomenologically speaking, the absence of physical embodiment (i.e., body) results in the absence of rich interaction between a person and her surroundings, by destroying our constant perception-action loop. However, as I stated earlier, philosophers and psychologists have been investigating the nature of this iterative process between perception and action, and from their works, we can discover opportunities to reconstruct our presence in virtual environments. Particularly, the following three embodiments of specific perception-action relations show the ways in which the virtual environments can extend our legitimate experience of being-in-the-world.

First, I pay special attention to **direct perceptual acts**, such as head turning and shifting, which in real world allow us to understand the invariant structure of 3D space from the variant “optical arrays”.²⁹ Coming from a phenomenological orientation, the *enactive* approach to perception is principally concerned with this idea, suggesting that our ability to perceive is constituted by our possession of bodily skills. Especially, Noë claims that we have access to environmental details as needed by turning and repositioning eyes and head.³⁰ The next section will describe how these acts can be mediated technologically into the 3D virtual setting.

Second, I also investigate our **locomotion with the mobile body**, as an extension of perceptual acts. I maintain Gibson’s position in which locomotion is treated as *proprioceptive*, as opposed to perceptive, since the movements and postures of the body are self-detected. While proprioception is often related to haptic feedback in HCI practices, I pay more attention to Gibson’s approach, since he places more emphasis on the visual attribute of proprioception. Referring to this special characteristic of locomotion, in the next section I will propose a novel method to achieve natural navigation in virtual 3D space.

Finally, I study **motor intentionality**, as a primary way to interact with objects in the world. Merleau-Ponty recognized this as something between movement as a third person process and thought as a representation of movement, and found that it makes ‘grasping’ or ‘touching’ different from ‘pointing’. I point out the current practice of HCI heavily relies on non-motor-intentional (i.e., cognitive) activities, whereas in real life that is not always the case. I will suggest the support of motor-intentional interaction with virtual content to improve the immersive sense of presence in virtual environments.

Overall, these three instances of perception-action loops will be the main themes of following chapters. I will demonstrate how philosophical and psychological insights could contribute to the natural experience of virtual environments, identify the bottlenecks found in the current interaction methods, and suggest how they can be overcome and improved.

²⁹ An optical array, or optical flow, is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer (an eye or a camera) and the scene; Gibson, Op.cit.

³⁰ Noë, A. (2004). *Action in Perception*. MIT Press, Cambridge, MA, 59.

CHAPTER 3 DIRECT PERCEPTUAL ACTS IN VIRTUAL SPACE

3.1 Perceptual Acts as Sensorimotor Skills

Modern psychologists and philosophers have been skeptical about our perceptual experience. In their view, perceptual consciousness is a kind of false consciousness; the visual world is a grand illusion. Daniel Dennett, among others, denied so-called snapshot conception of visual experience, according to which we perceive the picture-like world in sharp focus, uniform detail and high resolution, all at once. Dennett emphasized the fact that we have blind spots on our retinas, as well as bad parafoveal vision.¹ Thus he concluded that we are victims of an ‘illusion’ about the character of our own consciousness.

Alva Noë agreed on the fallacy of snapshot conception, but he also pointed out the problem of Dennett’s illusion hypothesis, in that we still can experience the world as richly detailed even though we lack internal representations of all that detail.² Noë solves this puzzle in two steps. First, our sense of the perceptual presence of the detailed world does not consist in our representation of all the detail in consciousness now. Rather, it consists in our ‘access’ now to all of the detail. That’s why, according to him, we still perceive a whole cat even though it is partly hidden behind the picket fence. Secondly, the basis of this access is our possession of sensorimotor skills, which is constitutive of sensory perception. One’s sense of the presence of the whole cat behind the fence can precisely consist in her implicit understanding by a movement of the eye, the head, or the body. Therefore, Noë suggests that an access to the detail of whole cat is controlled by patterns of sensorimotor dependence with which we are familiar; what we perceive is determined by what we do, which in turn is determined by what we are ready to do – we enact our perceptual experience. In this ‘enactive’ approach, as he calls, perception is a way of acting.³ Perceptual experience acquires content thanks to our possession of bodily skills. We take ourselves to be situated in an environment and to have access to environmental details as needed by turns of the eyes and head, and repositioning of the body. This view shares a great deal with Gibson’s thoughts on the visual perception as an access point for organisms to their environment, in which they act with their bodies:

One sees the environment not just with the eyes but with the eyes in the head on the shoulders of a body that gets about. We look at details with the eyes, but we also look around with the mobile head, and we go-and-look with the mobile body.⁴

The skillful movements of eyes and head, then, characterize perceiving as an active exploration of the environment. They are not powered by indirect inference, but direct bodily intention, as we have seen in the previous discussion. In contrast, our experiences with daily media such as texts, pictures, videos and websites, are mostly based on limited display spaces and/or indirect interaction modes, in which our skillful perceptual acts are neither encouraged nor supported. So far, current practice of virtual environments has inherited its form from such conventional media, and has often employed Cartesian analysis of perception and presence to engender immersive user experience; scientists and engineers try to create better ‘illusion’ by increasing the

¹ Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Co.

² Noë, A. (2002). Is the Visual World a Grand Illusion? In *Journal of Consciousness Studies* 9(5-6): 1-12.

³ Noë, A. (2004). *Action in Perception*. MIT Press, Cambridge, MA.

⁴ Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ.

perceptual quality and quantity (e.g., fidelity and scale of graphical representation) of mediation technologies, while often ignoring the intrinsic coupling between perception and action.

Nevertheless, our interaction with virtual environments, as extension of three-dimensional physical space, can greatly benefit from skillful bodily actions of perception. Particularly, head turning and shifting are primary modes of direct perceptual acts, constituting our spatial experience with surroundings.

First, head turning allows us to perceive scenes beyond the field of view (FOV) of our eyes. FOV refers to the maximum number of degrees of visual angle that can be seen instantaneously⁵. Without head turning, human vision covers approximately 140 degrees of binocular FOV with 40-45 degrees of peripheral vision in the horizontal direction.⁶ However, our foveal vision is much narrower and also accompanied by blind spots adjacently. Head turning is therefore a critical, not redundant, act to overcome such limitation in our vision.

Secondly, horizontal or vertical head shifting generates the motion parallax effect, providing spatial depth information. Motion parallax is one of the key elements that enable depth perception. When an observer moves her head laterally, the rate at which different objects are displaced in the retinal images depends on how far away those objects are from her.⁷ It demonstrates that we come to experience changes in our perspective as we move our body (parts), and that such variation allows us to understand the invariant three-dimensional structure of the space that surrounds us.

With the disembodied nature of current virtual environment practices, we lose this special perceptual capability that can be only motivated by our bodily actions. This critical aspect of perceptual experience cannot be digitally reconstructed by improving graphical reality or increasing display size/resolution, but by acknowledging and bridging the gap between perception and action in the interaction design.

3.2 System Supporting Direct Perceptual Acts

The use of head movements as interaction methods has been experimented in the domain of virtual reality (VR) technologies. Sutherland, when he invented the head-mounted display (HMD) more than four decades ago, already claimed:

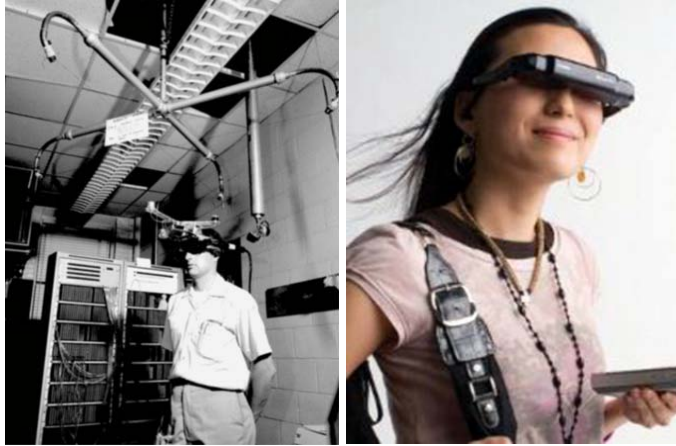
*The fundamental idea behind a three-dimensional display is to present the user with a perspective image which changes as he moves. (...) The image presented by the (...) display must change in exactly the way that the image of a real object would change for similar motions of the user's head.*⁸

⁵ Bowman, D. A., Kruijff, E., LaViola Jr., J. J., and Poupyrev, I. (2005). *3D User Interfaces: Theory and Practice*. Addison Wesley.

⁶ Rheingold, H. (1991). *Virtual Reality*. Touchstone, New York, NY.

⁷ Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA.

⁸ Sutherland, I. E. (1968). Head-Mounted Three Dimensional Display. In *Proceedings of Fall Joint Computer Conference*, 757-764.

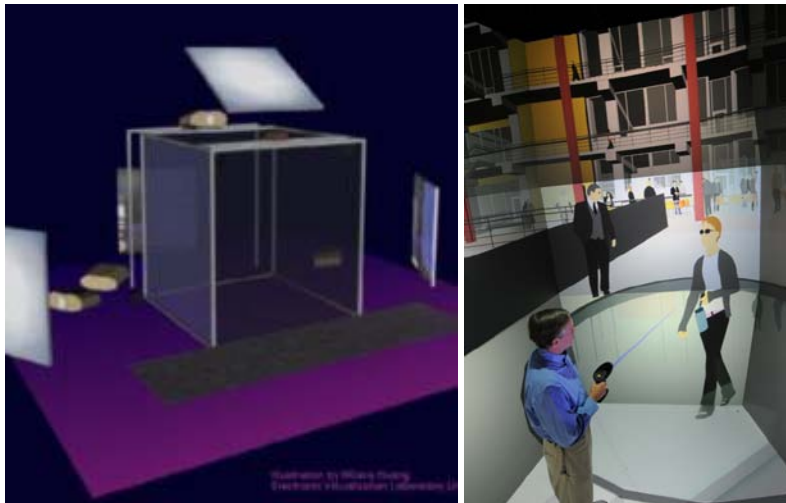


Left: the first HMD invented by Ivan Sutherland, in 1968.

Right: a personal display device developed by Vuzix Corp. in 2008.

Figure 3.1: Head-mounted displays (HMDs)

Since then, many different HMDs have been developed and used with tracking sensors, to generate the appropriate computer-generated imagery (CGI) for the angle-of-look at the particular time. This allows the user to "look around" a virtual environment by simply moving the head. However, even if the CGI placed in front of eyes block out the real world completely, HMDs still have limited FOVs (between 30 and 60 degrees horizontal), therefore impede a user's ability to acquire spatial information and develop spatial cognitive maps of unfamiliar spaces.⁹



Left: the original CAVE system developed at Univ. of Illinois at Chicago in 1991.

Right: StarCAVE, a five-sided VR room designed by Univ. of California at San Diego in 2008.

Figure 3.2: CAVE (CAVE Automatic Virtual Environment) systems

While HMDs suffer from narrow FOVs and motion sickness, surround-screen displays have been sought as less-intrusive ways to design a VR space. A surround-screen display is a visual output device that has three or more large projection-based display screens that surround the

⁹ Neale, D. (1998). Head Mounted Displays: Product Reviews and Related Design Considerations, Blacksburg, Dept. of Industrial Systems and Engineering, Virginia Tech, HCIL-98-02.

human participant.¹⁰ It was first developed at the University of Illinois at Chicago in 1991, and named CAVE (CAVE Automatic Virtual Environment). It consists of three walls and a floor, within which the viewer explores the virtual world by moving around.¹¹ There have been a number of variations on the design of surround-screen displays over the last decade, but they all commonly provide large FOVs, as well as the motion parallax effect when combined with a head tracking technology. Needing a large amount of physical space is often considered as one of their biggest disadvantages, and like HMDs, the user experience they provide is entirely isolated from the rest of physical surroundings.

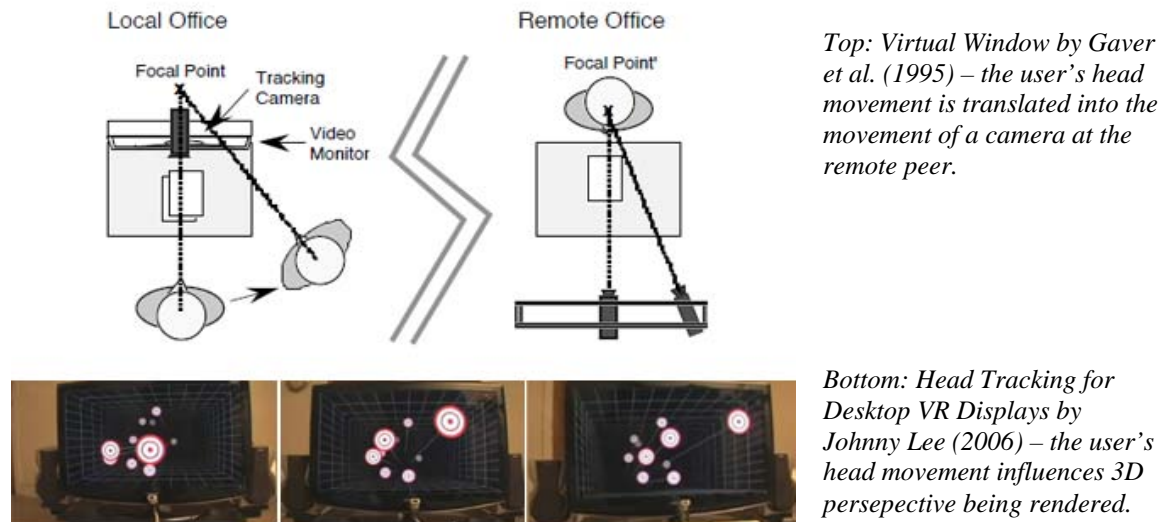


Figure 1.3: Systems supporting head movements

In parallel, support of perceptual acts in more personalized scales have been also studied. Gaver et al. observed that the typical arrangements of cameras and monitors in video-communication technologies provide only a fixed view of the remote location, outside the control of the observer.¹² Based on Gibson’s discussion of the “eyes in the head on the shoulders of a body,” their prototype video-communication system (called the Virtual Window, shown on the top of Figure 3.3) allowed users to explore a remote scene through head movements, using a head-tracking camera (locally) and a moving camera (remotely). This approach also utilized the user’s lateral and pivotal head movements to support motion parallax and gaze change. More recently, Lee demonstrated the motion parallax effect (on the bottom of Figure 3.3) in 3D virtual space, on a normal PC platform.¹³ His solution was simply based on an IR sensor-emitter on top of the display and a pair of retro-reflective markers attached on the user’s head. In both cases, however, narrow FOVs of personal displays still limited the overall user experience despite their active supports of head movements.

¹⁰ Bowman, D. A., Kruijff, E., LaViola Jr., J. J., and Poupyrev, I. (2005). *3D User Interfaces: Theory and Practice*. Addison Wesley.

¹¹ CAVE Automatic Virtual Environment. WWW URL: <http://www.evl.uic.edu/pape/CAVE>

¹² Gaver, W. W., Smets, G., and Overbeeke, K. (1995). A Virtual Window On Media Space. In Proceedings of the SIGCHI conference on Human factors in computing system. 257-264.

¹³ Johnny Lee. WWW URL: <http://www.johnnylee.net/projects/>

3.3 Design and Implementation of Interactive Non-Planar Display System

As we discussed above, prior works concerned with perceptual acts have mainly focused on 1) display space design or 2) interaction method development. However, their display spaces are either too narrow or inefficient, and the user experiences they afford are either isolated or intrusive. My design approach here is to find an optimal balance between natural user experience and spatial efficiency, so as to keep the perception-action loop sustainable. I propose a non-planar display design combined with natural interaction methods for direct perceptual acts.

3.3.1 Non-planar display space design

In the physical world, our field of view (FOV) is related to the way we are moving through the environment, and we have the opportunity to look around, and so build up a better picture of what is around us by exploration. In order to support such looking-around activity (i.e., head turning) in the virtual world, the proposed system uses multiple flat monitors that form a non-planar display space. This non-planar tiling of screens partially encloses the user's head, thus providing a wider FOV. When a display provides a wider FOV, more of the picture goes into the peripheral area of human vision.¹⁴ As the boundary between the screen and the rest of the environment is faded in the corners of our vision, the immersiveness of user experience is getting enhanced.

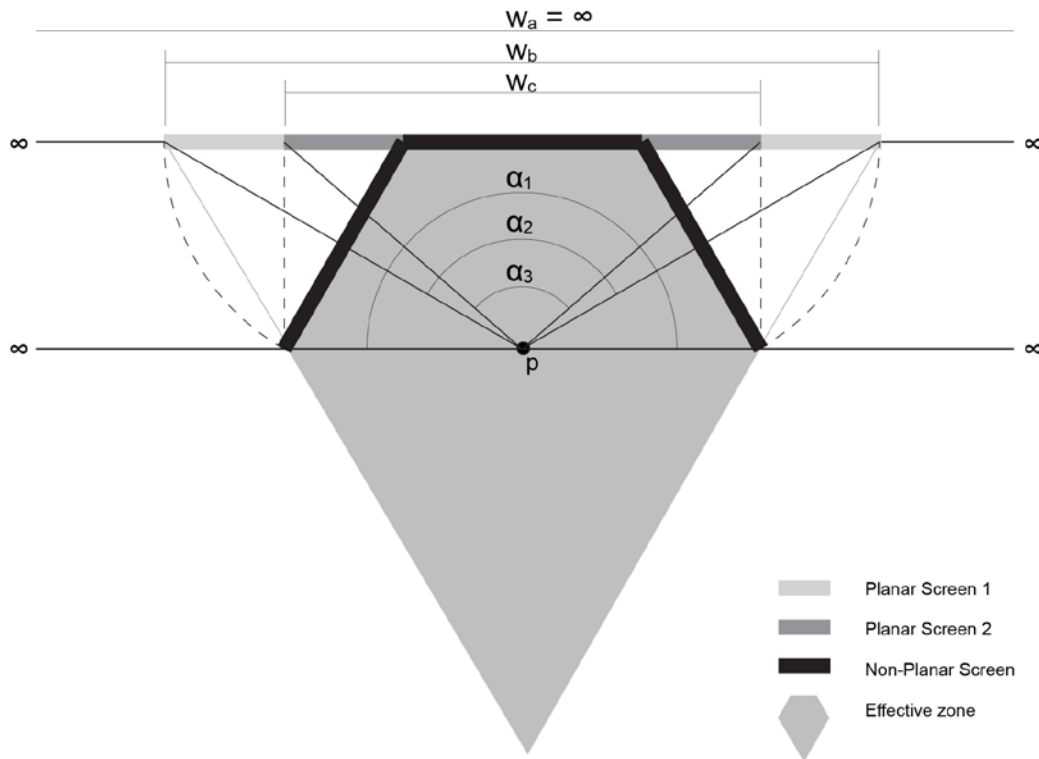


Figure 3.4: FOVs in non-planar and planar display spaces

¹⁴ Reeves, B. and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications, New York, NY.

Assuming that there is a 20-inch widescreen LCD display viewed at 50 cm (these are typical numbers for desktop display size and viewing distance), the viewer has approximately 49 degrees of horizontal FOV. If the user wants to increase the FOV up to 180 degrees (which is the total horizontal FOV of normal human vision¹⁵) while preserving viewing position in relation to the display, the viewer will need a non-planar (or curved) display space, otherwise only a flat display wall with infinite horizontal dimension could provide a maximum FOV of 180 degrees.

Figure 3.4 illustrates why this design approach is compelling in terms of FOV, combined with a highly-efficient use of physical space. Given that the user's head is positioned at point p , without requiring an unlimited wall space (w_a), **Non-Planar Screen's** FOV α_1 is wider than:

- α_2 (from **Planar Screen 1** whose width is equivalent to **Non-Planar Screen's** unfolded length w_b), and,
- α_3 (from **Planar Screen 2** whose width is equivalent to **Non-Planar Screen's** projected length w_c).

And as far as the user's head moves around within **Effective Zone**, this is always true.



Figure 3.5: A prototype vertical non-planar display with three monitors

As a proof-of-concept configuration, I arranged three flat screens in upright position with 120-degree angle between one another. Figure 3.5 shows the prototype display consisting of a 15.6-inch laptop computer and two 15.6-inch LCD displays. The video-out signal from the laptop computer is transferred to two LCD monitors through a VGA signal splitter¹⁶, and viewports for three displays are arranged in order by using OpenGL graphic library. In this non-planar display space, I track the user's face position, which then is translated into the location of a virtual camera cluster. The system uses a face tracking function implemented as a

¹⁵ Rheingold, Op.cit.

¹⁶ Matrox Electronic Systems Ltd. WWW URL: <http://www.matrox.com>

vision process per each frame of a webcam feed, and renders the 3D scene on the non-planar display space by deforming view frustums asymmetrically. Details will be described in the following section.

3.3.2 Motion parallax and view-dependent rendering (VDR)

Using the above mentioned display space, the proposed system provides visual depth information arisen from motion parallax, by tracking the user's face position in three-dimensional space and applying it to the calculation of perspective projection in 3D graphic rendering pipeline. This is commonly referred to as view-dependent rendering (VDR). Figure 3.6 demonstrates the motion parallax effect generated by VDR. As the user's head shifts from left to right, the relationships between objects in the scene keep changing. While the smaller (red) sphere barely moves in relation the screen, the (white) teapot moves to the right at high speed and eventually disappears. The relative distances from the user's viewpoint to the smaller (red) and larger (blue) spheres are gradually discovered as perspective changes over time. These events provide the user with visual depth cues, allowing her to realize that the smaller (red) sphere is closest object, the (white) teapot is farthest one, and the larger (blue) sphere is positioned in between.

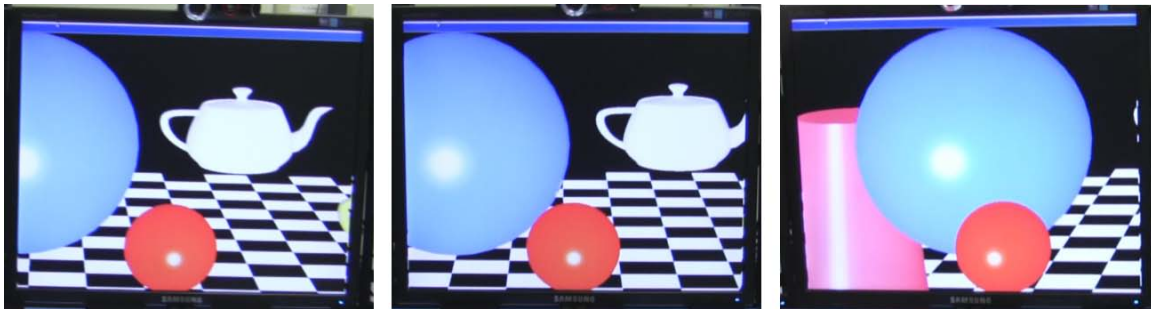


Figure 3.6: View-dependent rendering and motion parallax

3.3.2.1 Face tracking with OpenCV

The proposed system tracks the user's face position in the X, Y, and Z directions (3DOF). Positioned at the top-middle location of the central screen, a webcam delivers a live video feed to the system, at 30 fps (frames per second) in QVGA (320x240) resolution. After being grayscale, each video frame is scanned for detecting a face. I use a rejection cascade built from the Viola-Jones classifier¹⁷, which comes with the OpenCV library¹⁸. This cascade was pre-trained with around 9,000 face images, tested with another 1,000 images, and stored in a XML file. A rejection cascade is ideal for a real-time process because it declares true class detection only if the computation is made through the entire cascade, otherwise terminates the computation at any stage of each node where "no face" is declared¹⁹.

¹⁷ Viola, P., and Jones, M. J. (2004). Rapid Object Detection Using a Boosted Cascade of Simple Features. In *Proceedings of Computer Vision and Pattern Recognition*, 511-518.

¹⁸ OpenCV: Open Computer Vision Library. WWW URL: <http://www.opencv.org>

¹⁹ Bradski, G., and Kaehler, A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media.

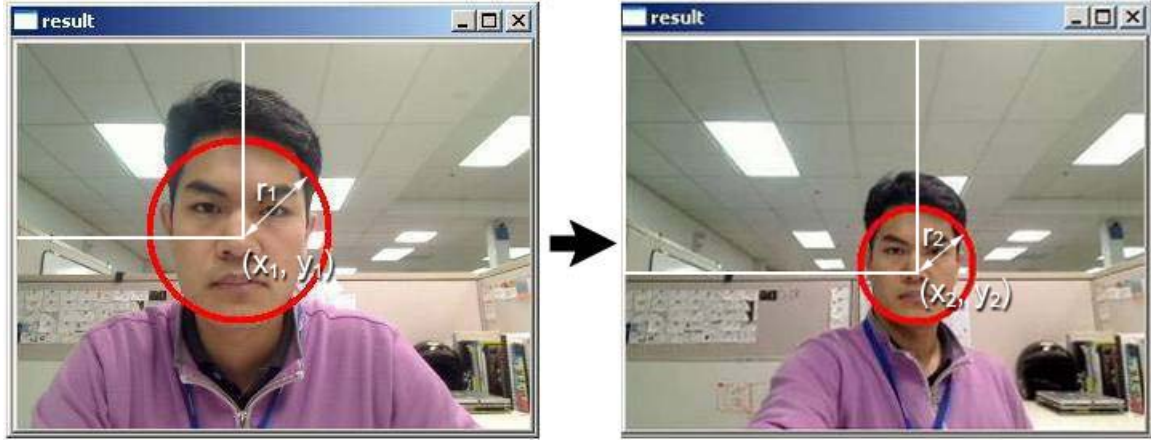


Figure 3.7: Face tracking test with OpenCV library

For each frame, the proposed system retrieves screen coordinates and size (radius) of a found face from this face detector, as shown in Figure 3.7. When the face is positioned at the center of screen (i.e., screen coordinates (160, 120)) and maintains the default size²⁰, the virtual camera is located at (0, 0, 0) in the 3D model space. Otherwise the face position (x, y) in the size of r is translated into the 3D coordinates:

$$(c_1(x - 160), c_1(120 - y), c_2(r_d - r))$$

, where c_1 and c_2 are calibration factors and r_d is the default size of the face.

3.3.2.2 Asymmetric view frustum cluster

Rendering 3D content on non-planar display spaces is non-trivial, since it requires multiple view frustums. The proposed system applies a separate view frustum for each planar screen component; otherwise a considerable amount of perspective distortion will be introduced. Figure 3.8 shows a pair of rendering samples, which were generated with a single view frustum (top) and multiple view frustums (bottom), respectively. In order to cover the entire cubic space (grid), FOVs have to be maximized in both cases; however the single frustum with an extremely-wide FOV results in significant linear distortion on sides, compared to the multiple view frustums.

In the proposed system, all frustums together form a cluster that has a fanwise shape, and images on their view planes are seamlessly connected to one another. Unlike typical 3D rendering pipelines where the view-plane normal is assumed to pass through the center point of the screen surface, the view-dependent rendering technique needs to move the normal vector from the center of screen. This results in asymmetric deformation of standard view frustum shape. Based on the physical dimensions of display surfaces and the user's face position, the proposed system calculates, per video frame, the coordinates of left, right, and near clipping planes for each view frustum. As depicted in Figure 3.9, clustering of asymmetric view frustums enables seamless rendering of 3D content across multiple display elements, with correct perspective projection.

²⁰ The default size of a face can be determined by positioning the face in front of the central vertical display surface at the distance of h (the height of each display component), as illustrated in Figure 11.

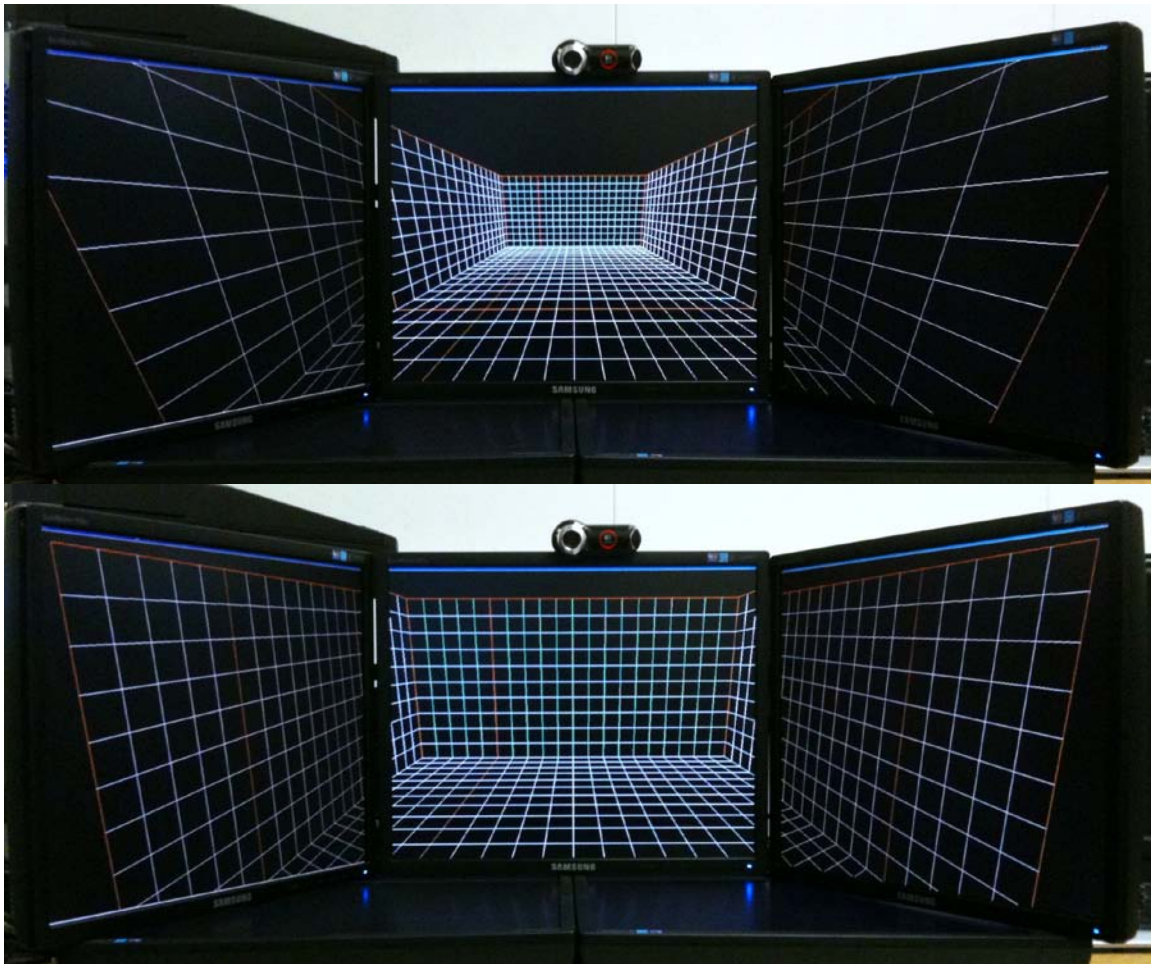


Figure 3.8: Comparison between single (top) and multiple (bottom) view frustums

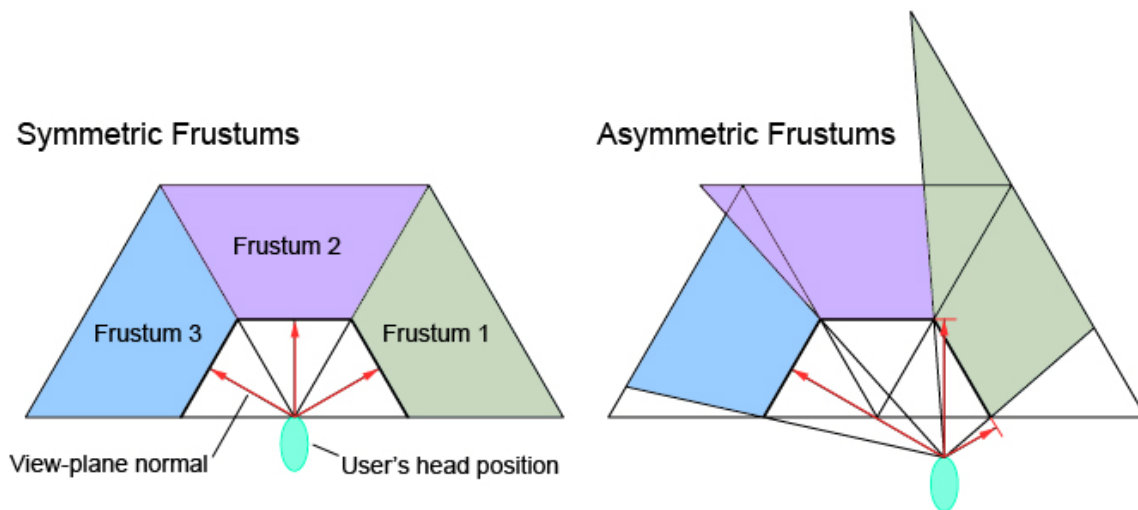


Figure 3.9 Symmetric vs. asymmetric view frustums

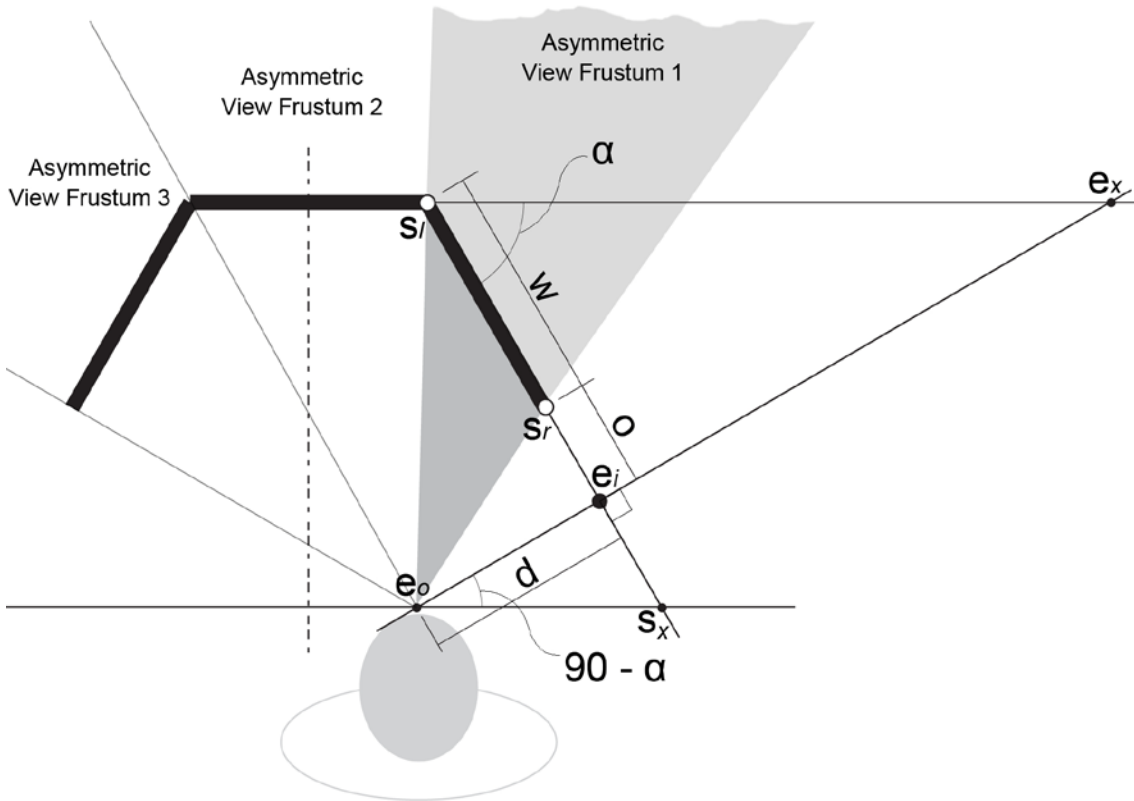


Figure 3.10: Calculation of asymmetric view frustum in a non-planar display space

The correct shapes of view frustums in a vertical non-planar display space, as illustrated in Figure 3.10, can be calculated from:

- the angle (α) between planar display components,
- the width (w) of the planar display component,
- the intersection point (e_i) at which the normal vector from the head position (e_o) meets the infinite plane in which the planar display component lies
- the distance (o) between the above normal vector and the right corner point (s_r) of the planar display component, and

the distance (d) between the intersection point (e_i) and the head position (e_o).

For example, the asymmetric view frustum 1, which corresponds to the user's current head position, can be determined through the following procedure:

- 1) calculate the coordinates of e_x and s_x (given that α is always known),
- 2) calculate the coordinates of e_i from the line $e_o e_x$ and the line $s_l s_x$, which are always guaranteed to intersect each other,
- 3) calculate o and d from e_i , and
- 4) finally, determine the shape of view frustum with $(w/2 + o)$ as the horizontal offset of left and right clipping planes, and d as the distance of near clipping plane from the virtual camera. The vertical offset of top and bottom clipping planes is simply determined as the opposite of (vertical) head movement.

In the same manner, I build the view frustum 3 for the left planar display component. Meanwhile, the view frustum 2, for the center planar display component, can be easily determined by

applying the opposite of head movement in X, Y and Z directions, to left/right clipping planes, top/bottom clipping planes and the near clipping planes, respectively.

3.3.3 Horizontal extension of non-planar display space

The proposed vertical non-planar display space efficiently extends the visual field in the horizontal dimension, but not in the vertical one. With typical layouts of our daily desktop settings in consideration, adding horizontal screens to the proposed non-planar display design can improve the natural experience with virtual environments. As an ideal configuration, I use five identical LCD monitors (by adding two more to the three-monitor setup) to build this extended display space. In the top of Figure 3.11, vertical components maintain 120 degrees one another, and two horizontal components fit into the existing setup. The bottom of Figure 3.11 demonstrates that it also requires asymmetrical formation of view frustums for horizontal planar display components, and their shapes depend on the near clipping plane of **view frustum 2**, as well as the user's head position \mathbf{e}_o in relation $(\mathbf{f}_x, \mathbf{f}_y, \mathbf{f}_z)$ to the display space. This results in a volumetric display that is curved horizontally and vertically, extending the user's FOV in both dimensions.

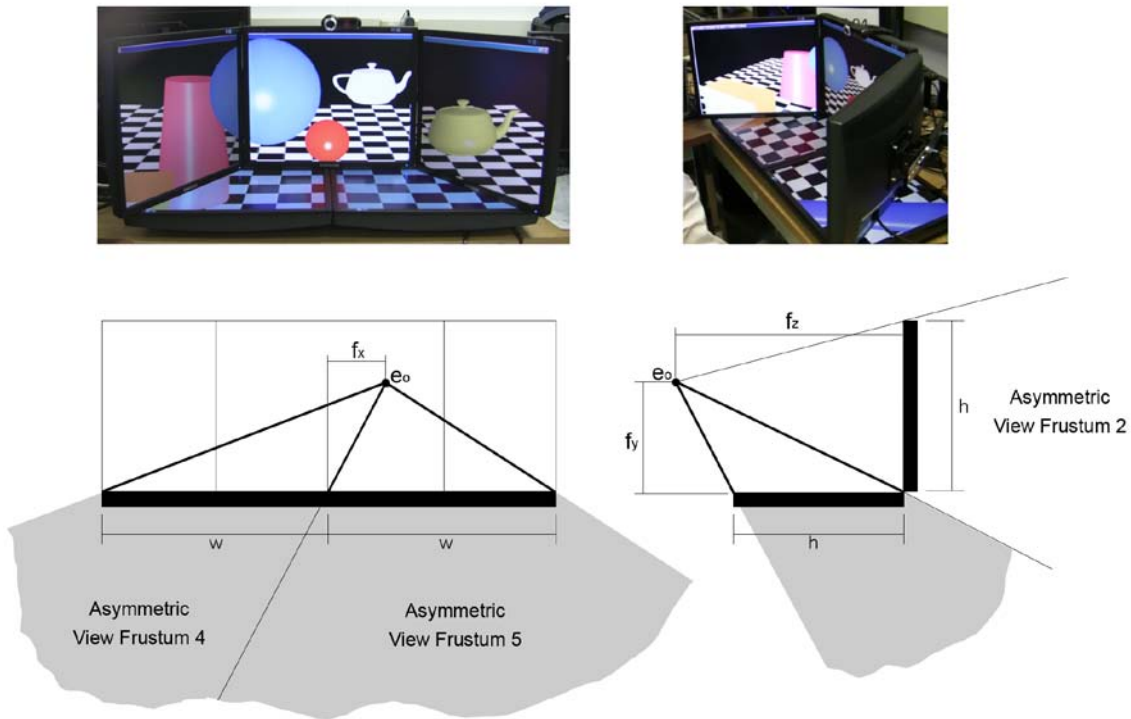


Figure 3.11: Horizontal extension of display space and view frustums

3.4 Discussion: VDR and Stereoscopic Displays

One of the advantages the proposed system and method provide is the support of motion parallax, which allows us to perceive visual depth information even without any stereoscopic technologies. Motion parallax is based on retinal images that are displaced over time, whereas stereoscopic vision is based on pairs of simultaneous retinal images. Therefore a motion parallax effect can be successfully simulated with a normal (2D) screen by displaying images sequentially, while

stereoscopic vision can only be simulated with specific techniques such as anaglyphic glasses, HMDs, parallax barrier displays, or lenticular displays.

Most stereoscopic vision techniques may cause user fatigue, because they do not consider the conflict caused by the difference between visual information from the eyes and vestibular cues from the inner ear. In the proposed system, however, the user will not experience such fatigue or dizziness, because I simulate the motion parallax effect with 3D content by linearly translating the perspective based on the user's head position.

In addition, stereoscopic displays without the support of motion parallax provide us contradictory depth information when we move our head around. This is due to the fact that binocular disparity is originated from the flat imagery, not from the actual three-dimensional space. As we move our head laterally in front of stereoscopic content, the direction of depth is always following our viewpoint, thus generates an illusion in which farther objects move faster than closer objects toward the opposite direction of head movement. This is exactly the reverse of normal motion parallax we would experience in the physical environment. In this sense, the method I proposed here actually compensates the defect of stereoscopic technologies, rather than competing with them. While the integration of VDR and stereoscopic displays have been experimented with HMDs or large-scale surrounding displays such as CAVE, the importance of action-based perception has been ignored in the emerging trend of 3D televisions, games and virtual worlds.

CHAPTER 4 INTERACTIVE CHAIR: PROPRIOCEPTIVE LOCOMOTION CONTROL

4.1 Proprioception and Locomotion

We have seen that perception is not separable from action, and have demonstrated ways in which our direct perceptual acts play important roles in the experience of both physical and virtual environments. However, there still remains a question one may ask: if we already seem to enjoy the virtual environment thanks to our skillful use of keyboard, mouse or gamepad, exploring the 3D space without moving our head or body necessarily, just rotating or shifting the imagery on the screen instead, then what is wrong with this convenient way of perceiving the world? Isn't this what the virtual world is for after all?

As a direct response, psychologists answer that question with scientific experiments. Among others, Crowell and his colleagues demonstrated that extra-retinal information is critical in the interpretation of visual input during self-motion¹. In their experiment, subjects perceived the simulated self-motion accurately while smoothly shifting the gaze by turning the head, but not when the same gaze shift was simulated in the display. They found that accurate perception during active head turns is mediated by contributions from three extra-retinal cues: vestibular canal stimulation, neck proprioception and an efference copy² of the motor command to turn the head. With current modes of interaction in virtual environments, we experience changes in retinal information without updating the other cues, which normally accompany visual input in the physical environment. Head turning and shifting have to be considered in 3D interaction design for this reason.

However, our direct perceptual acts in the physical world do not end with head movements; as living organisms we also use the rest of body to interact with surroundings, and ultimately move around with the whole body in space, namely, locomotion. When we perceive the world in this mode, we do not only experience the environment surrounding our body, but also experience the body itself. Gibson considers this particular property as a kind of proprioception, and describes the case of locomotion as an example:

(...) the motion of one's body relative to the stationary environment, whether active or passive, can be detected by vision, and this is a case of proprioception. Locomotion, as distinguished from object motion, is specified by transformation of the ambient optic array as a whole. The movements and postures of the body are detected (in several independent ways) whether they are imposed by outside forces or are obtained by an action of the observer himself.³

What we lose, when we move around in virtual environments, is then the ability to relate our body and its movements to the space that surrounds us virtually. Instead, we are forced to develop completely new skill sets to manipulate hardware interface, so as to simulate our locomotion independent of proprioceptive cues immanent in our body. Hence, I suggest that the connection between proprioception and locomotion is another instance of perception (of body) – action (of moving around) loop that is easily broken in virtual environments.

¹ Crowell, J. A., Banks, M. S., Shenoy, K. V., and Andersen, R. A. (1998). Visual self-motion perception during head turns. In *Nature Neuroscience*, 1, 732-737.

² *Efference copy*: a signal from a part of the brain involved in the control of eye movements describing the motor command to turn the eye; Howard, I. P. (1982). *Human Visual Orientation*. John Wiley, Chichester.

³ Gibson, J. J. (1972). A theory of Direct Visual Perception, In Royce, J. and Rozenboom, W. (Eds.). *The Psychology of Knowing*. Gordon & Breach, NY, 215-240.

4.2 Human Body and Spatial Concepts

Bodily movements help us understand space in reference to our body. Our head and feet unite to form a principal (vertical) body-axis, which is normally aligned with gravity. Our body also has a front and a back (sagittal). There is also a third body-axis, left and right (horizontal), which is resting on a less clear distinction. People conceptualize space in terms of these three axes and their relation to gravity.⁴ The nature of our bodies, with their three axes as well as our perceptual and motor skills, determines how we define spatial relations in the physical world. Franklin et al.⁵ devised the ‘spatial framework analysis’ to describe the mental models which readers derive from narratives. When subjects read a story describing a scene, they mentally create a spatial framework consisting of three body axes, and associate described objects to that framework. For an upright observer, in this mental model, the head/feet axis is most accessible because it is physically asymmetric and correlated with the fixed environmental axis of gravity. The front/back axis is the next most accessible since it separates the world that can be seen and manipulated from the world that cannot be easily perceived or manipulated. The left/right axis is least accessible because it has no salient asymmetries. While reading a description, the subject adopts the perspective of the person in the narrative and mentally occupies the position of that person in the scene. Therefore, it was concluded that the subject’s own body is at the center stage in human spatial concepts.

A similar analysis was applied to the case of perception. Instead of reading a description, Jolicoeur and colleagues asked the subjects to ‘perceive’ a person and locate objects relative to that person.⁶ Again, head/feet axis was identified the fastest. However, it differs from the spatial framework: it does not require observers to place themselves mentally in another person’s perspective or create a mental spatial framework. Instead, observers extract the directions by analyzing the person’s intrinsic axes, then visually scanning in the appropriate direction. It eliminates any conflict between the subject’s actual viewpoint and that of the other person. This approach has been known as ‘intrinsic computation analysis.’

On these bases, Mylov investigated our spatial concepts in VR.⁷ According to him, the avatars, the representation of ourselves, may help us to establish a frame of reference by their geometric (intrinsic) features. Appearing as objects on the screen, the avatars invite an intrinsic analysis meaning that the head/feet axis is the fastest computed, even if they are inclined. The front is also the salient axial direction with the backspace, as it establishes the front of the object as the side facing the observer.

⁴ Bryant, D. J. (1998). Human spatial concepts reflect regularities of the physical world and human body. In *Representation and Processing of Spatial expressions*. Lawrence Erlbaum Associates Inc. pp. 215-230.

⁵ Franklin, N., and Tversky, B. (1990). Searching imagined environments. In *Journal of Experimental Psychology: General*, 119, 63-76.

⁶ Jolicoeur, P., Ingleton, M., Bartram, L., and Booth, K. S. (1993). Top-bottom and front-behind decisions on rotated objects. In *Canadian Journal of Experimental Psychology*, 47 (4): 657-677.

⁷ Mylov, P. (2002). On Space, Its Time, and Spatiotemporal expressions. In *Virtual Space: spatiality in Virtual Inhabited 3D Worlds*, Springer-Verlag, 47-70.

4.3 Technologies for Locomotion Control in 3D Space

Many efforts have been made to develop virtual locomotion interfaces based on our bodily motions. The most direct and natural technique for locomotion in 3D space is, obviously, to physically walk through it. The HiBall tracking system⁸ optically tracks a wide area by using a scalable tracking grid on the ceiling. This system uses user-mounted optical sensors to compute the user's position and orientation. However, this method is spatially limited since the size of the environment must be less than tracking range. The space also needs to be free of obstacles.

Alternatively, walking simulation has been used to implement interfaces for locomotion in 3D space. Noma and Miyasato⁹ used a common treadmill to provide a walking motion and feel without translating the users' body. Since the basic treadmill does not allow the user to turn naturally, they also tracked the user's head and feet to detect when the user is trying to make a turn, by analyzing the direction the feet are pointing, deviation of a foot's motion from the forward direction, and other factors. While this approach removes the limitation on the size of the environment, it still suffers from problems including user fatigue and too much latency for sudden or sharp turns. In another approach, CyberCarpet¹⁰ proposed an omni-directional treadmill that allows the user to perform locomotive motion in any direction. Moreover, Nurulla and Ray Latypov developed VirtuSphere¹¹, an enclosed VR device in which a user can walk in place while rotating the spherical shell in any direction.

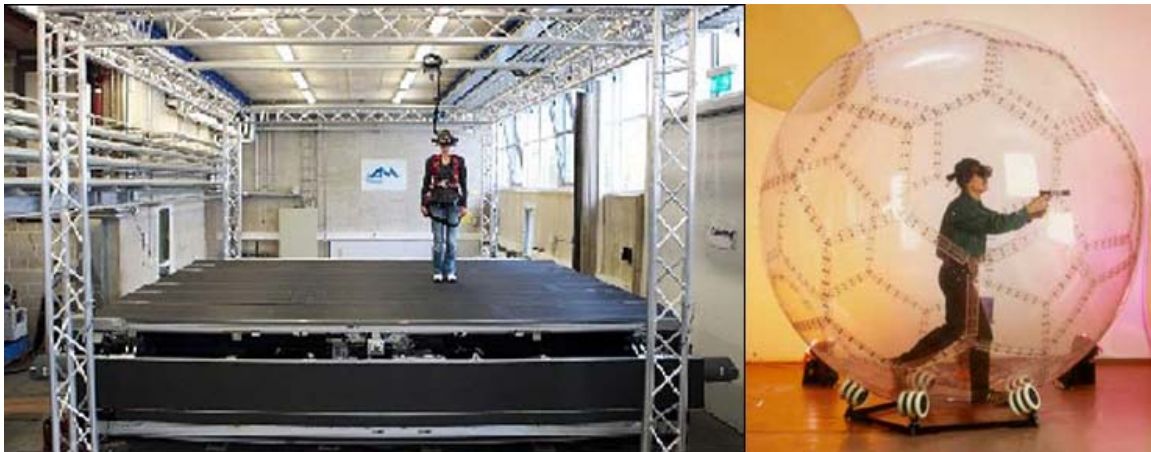


Figure 4.1: CyberCarpet (left) and VirtuSphere (right)

However, none of these approaches are free from the problem of user fatigue. Researchers empirically realized that users have much lower tolerance for physical motion in a virtual environment than they do in the real world. The gaze-directed steering is the most common technique to solve this problem. It allows the user to move in the direction toward which he/she

⁸ Templeman, J. N., Denbrook, P. S., and Sibert, L. E. (1999). Virtual Locomotion: Walking in Place through Virtual Environments. In *Presence: Teleoperators and Virtual Environments*, 8(6): 598-517.

⁹ Noma, H. and Miyasato, T. (1998). Design for Locomotion Interface in a Large Scale Virtual Environment-ATLAS: ATR Locomotion Interface for Active Self-Motion. In *Proceedings of the ASME Dynamic Systems and Control Division*, 111-118.

¹⁰ CyberWalk Project. WWW URL: <http://www.cyberwalk-project.org/>

¹¹ VirtuSphere. WWW URL: <http://www.virtusphere.com/>

is looking. The gaze direction is usually obtained from the orientation of a head tracker. As an example, *TrackIR*¹² implements this idea by using IR camera and retro-reflective markers. This gaze-directed steering (i.e. head tracking) is easy to understand and control, however its usage limits our natural body movement by rigidly coupling our gaze and body in the virtual space. Since our head movement originally serves to direct perceptual functions, as we have discussed previously, it is not natural to couple the gaze direction with the locomotive direction.

4.4 Design and Implementation of Interactive Chair

Interestingly, prior works on virtual locomotion techniques suffer from the similar problems we have already noticed in the previous chapters – unnatural user experience and spatial inefficiency. I propose to utilize our proprioceptive sense of bodily movements and innate spatial concepts in the design of virtual locomotion control. I argue that it does not have to involve the same amount of physical efforts that are necessary for locomotion in the physical space, but it can be still achieved even with sitting on a chair. After all, chairs are the most significant objects in our daily environment as we spend most of time sitting on one of them, and in practice, our experience with virtual environments are mainly performed while sitting on a chair, too. The proposed system, named Interactive Chair, particularly detects subtle transitional sitting postures of the torso that are directly related to two types of the proprioceptive sense: yawing (rotation along the vertical body axis) and leaning (tilting upper body).

4.4.1 Spatial correlation between sitting postures and 3D locomotion

Position and orientation of torso relative to thighs determine most of sitting postures. Especially, torso movements generate varied forms of sitting postures, reflecting the occupant's bodily intention. This allows the Interactive Chair to leverage the spatial correlation between dynamic sitting postures and locomotive motions. I focus on two types of sitting behaviors – leaning and swiveling actions, in the next section.

4.4.1.1 Leaning behaviors and locomotion

In typical office settings, a person sitting on a chair often leans over the desk to pay special attention on the content of screen or document. To the contrary, the occupant leans against the backrest of chair to relax or relieve attention. In the case of standing or walking behaviors, this forward-leaning motion usually initiates, therefore is followed by, the action of “approaching” or “running”. In the same analogy, the back-leaning motion (including the transition from forward-leaning position to upright position) initiates “backing off” or “stopping” action. The proposed system leverages this spatial conception that correlates our sitting posture with locomotion. It provides the user with a direct traveling control, with which the 3D perspective stays in place when the user's torso is in upright position; otherwise the perspective (and the associated virtual camera) should move corresponding to the current leaning direction of user. The system also uses the degree of torso inclination to control the locomotive speed in the current leaning direction. It allows the user to accelerate or decelerate the virtual locomotion in real time.

¹² NaturalPoint, Inc. WWW URL: <http://www.trackir.com/>

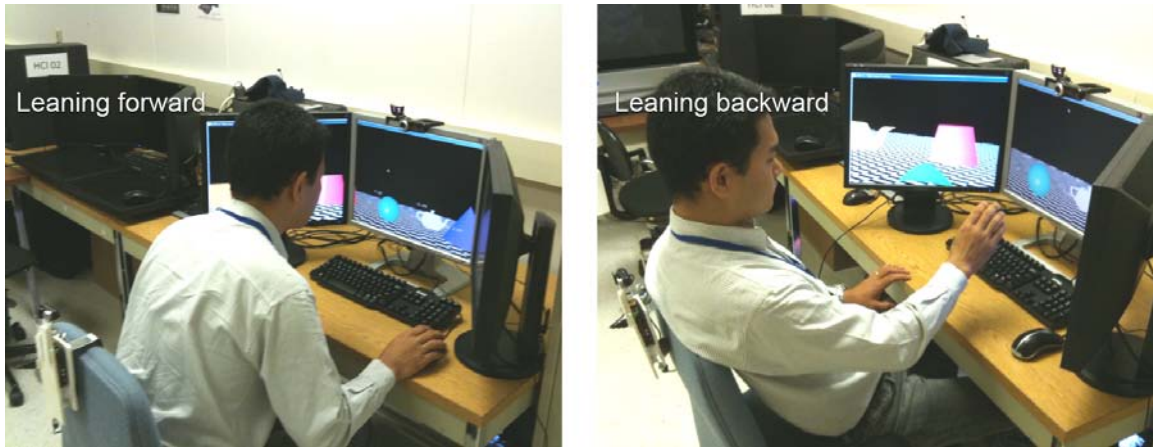


Figure 4.2: Leaning forward/backward (sagittal plane) with Interactive Chair

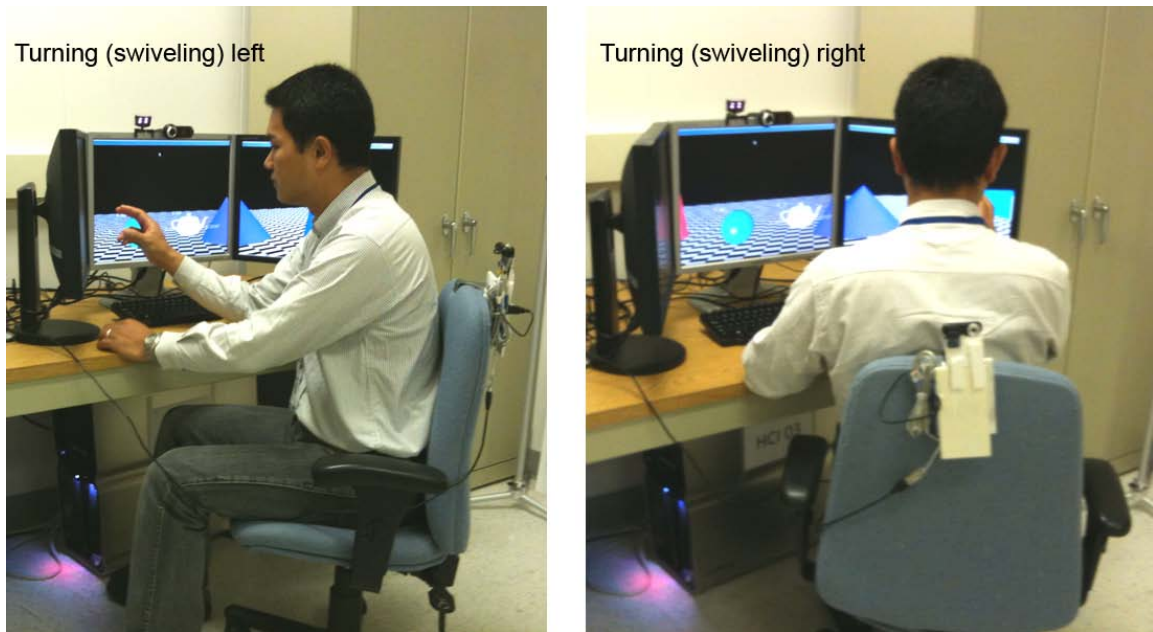


Figure 4.3: Swiveling with Interactive Chair

4.4.1.2 Swiveling and orientation

Torso movements on a swivel chair naturally involve turning motions. Originally, the swiveling function of a chair is to give a wide angle of spatial access to the user being seated. For the locomotion control in 3D space, this swiveling motion on the chair can be translated as the rotation of entire body orientation. As the user turns the swivel chair from the feet, the camera view in the 3D space follows the swiveling direction. For example, turning the chair toward right will rotate the virtual camera clockwise (from the top view). This is in line with the current practice of virtual environments, in which the camera rotates toward the direction of mouse movement. In the real world analogy, the swiveling motion on the chair can be understood as the action of turning direction while walking.

4.5 Sensing Transitional Sitting Postures

Using a chair as a computer interface is not a novel approach. Tan et al.¹³ developed *SensingChair* that supports multimodal interaction through the user's torso movement. The main idea of their prototype was to develop a real-time system that feels its occupant through a layer of 'artificial skin'. Two pressure sensor sheets, made possible with commercial pressure distribution measurement system, were surface-mounted on the seat and the back rest of an office chair to capture various sitting postures (Figure 4.4).

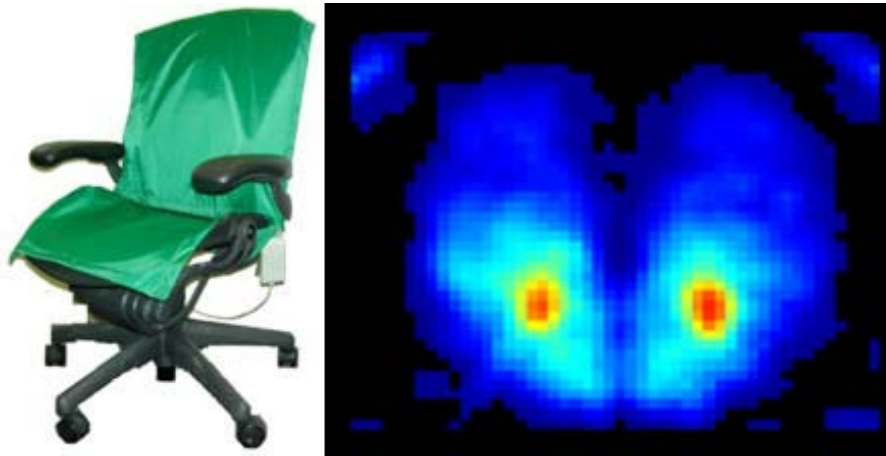


Figure 4.4: SensingChair and sitting pattern image

Since the analysis of pressure distribution in this system required heavy computation process for pattern recognition, it was only suitable for static applications such as the force deployment control of vehicle airbag, or the maintenance of healthy postures of users. Tracking transitional postures in real time was not technically feasible. In contrast, the proposed system detects transitional sitting postures, by utilizing the following sensing methods on a typical swivel chair.

1) Detecting leaning motion:

Distance between the user's back and the backrest of the chair was measured by using an ultrasonic (40 kHz) range finder (*Ping* by Parallax Inc.), which computes the distance based on time of flight (TOF) of brief chirps. This sensor was installed above the backrest by using a height-adjustable mount made of foam board.

2) Measuring swivel motion

A standard single-turn potentiometer (10K Ohm) was used to detect the swivel angle of the chair as rotated by the user. In order to increase the sensitivity of sensor by 5 times, the chair shaft was enhanced with a cylinder-shaped bucket (diameter: 10 inches), so it could engage with a 2-inch wheel that is fixed to the potentiometer head.

¹³ Tan, H. Z., Slivovsky, L. A., and Pentland, A. (2001). A Sensing Chair Using Pressure Distribution Sensors. In *IEEE/ASME Transactions on Mechatronics*. *IEEE/ASME*, 6(3): 261-268.

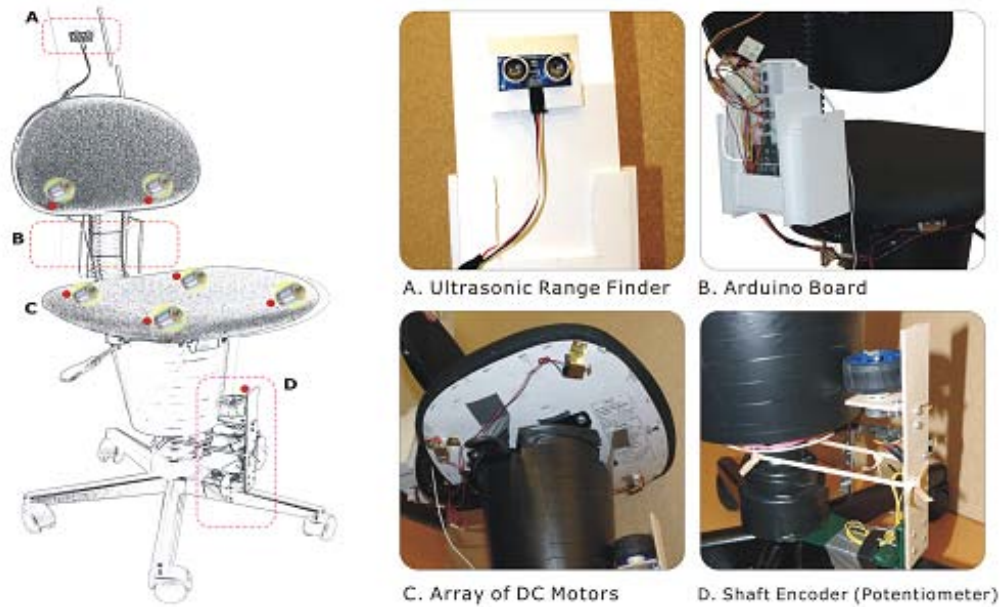


Figure 4.5: Prototype Interactive Chair and sensor systems

4.5.1 Sensor arrangement and signal processing

The ultrasonic range finder measures the distance to a target object up to 3 meters. The output value ranges from 0 to 4096. It sends an ultrasound tone and waits for an echo, the width of which determines the distance to the object. In this study, the sensor detected the distance in every 20 millisecond (50 Hz), which should be frequent enough for controlling the locomotive speed of 3D avatar in real time. Assuming that the actual distance between the user and the backrest would be no longer than 2 feet (about 0.6 m) mostly, the output signals were windowed between 0030 and 0300, in order to filter out potential noise data in advance.

The single-turn potentiometer detects the rotational movement up to 270 degree, and its output value ranges from 0 to 127. Given that the diameter ratio between the potentiometer and the chair shaft is 1:5, the output value increases by 1 every time the chair turns 0.4 degree clockwise (approximately). Accordingly, the user only needs to rotate the chair up to 27 degrees either right or left side to reach the maximum range of sensor. This makes it easy to control gaze or body orientation in the virtual space, minimizing the fatigue of users.

Sensor signals are controlled and collected by a microcontroller *Arduino* board, which then transferred data to PC. For simulating 3D space, *Torque* game engine was chosen as the software platform. The engine's source code was modified to implement a serial port (COM) based communication channel, which connects to *Arduino* board. In order to minimize the system noise level, 5-point triangular smoothing algorithm was applied to sensor signals. Given an output value x_i , a weighted smoothing function $f_s(x)$ is calculated as $f_s(x) = (x_{i-2} + 2x_{i-1} + 3x_i + 2x_{i+1} + x_{i+2}) / 9$.

4.5.2 Data mapping in the 3D engine

Data from the ultrasonic range finder were translated into a simple curve ($U' = 1.5 * Su^2 - 30$, where Su = signal value from the ultrasonic range finder), so that the forwarding speed of locomotion can increase quadratically as the user's torso leans further away from the backrest. The forwarding speed ranged from 0 to 2.0.

The output from the potentiometer can be translated in two different ways. On one hand, when the gaze needs to be coupled with the body orientation in the 3D space, the potentiometer linearly increases or decreases the angular rate (yaw) of virtual camera view associated with the avatar ($yaw = 0.8 * Sp$, where Sp = signal value from the potentiometer). On the other hand, when the gaze needs to be decoupled from the body orientation, each signal value represents an absolute angular position of the chair shaft. This let the user easily return back to the original gaze direction after looking around, by just reorienting the chair forward. In this project, the user could activate the decoupled gaze control by simply holding down the shift key.

4.6 Swiveling Motions and Optical Flow

The proposed system exploits the swiveling function of a chair as a part of the locomotion control, since it is the most direct and natural way to align the user's physical orientation with the forward direction in the virtual space. However, the proposed prototype system is based on a hardwired integration between sensor modules and a chair, thus become impractical to be used. In the second version of prototype, for this reason, I optimized the sensor module into a portable form factor by employing a simple computer vision process. All sensing hardware, such as an ultrasonic range finder (sonar), a micro-controller (*Arduino* board), a webcam, and a USB hub, were mounted in a small box and attached to the back of the chair. The box is connected with a single USB cable to the rest of the system.



Figure 4.6: Portable sensor box configuration

In this version of prototype, the system assesses motion between two video frames from a webcam attached to the backrest of chair, instead of detecting the rotation of chair shaft. By tracking and comparing a set of features in those two consecutive video frames, the system determines a motion vector, i.e., optical flow, for each feature. Then it calculates the mean of optical flows from all features, so as to provide the final direction and rotational velocity of the chair. This process is a kind of ego-motion detection, as the user's torso rotates together with the device, in this case, both the chair and the webcam (imager). Since yawing motion generates strong optical flow fields compared to any type of shifting motions, the system is able to detect the swiveling motion with high precision in real-time. I used OpenCV library to extract features from each video frame, and then computed the overall optical flow vector as depicted in Figure 4.7.

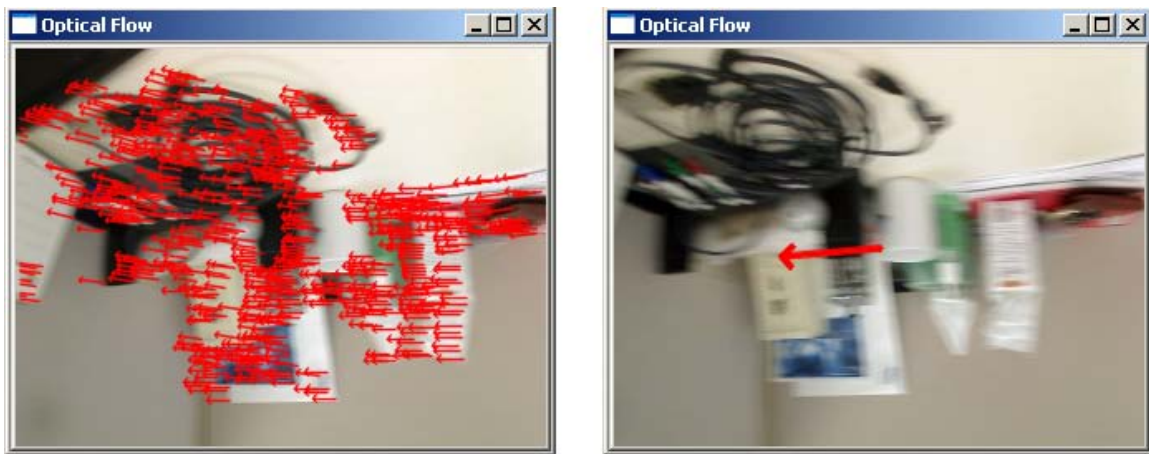


Figure 4.7: Feature extraction and optical flow vector

Finally, the system translates this optical flow vector into the direction and speed of the rotation (yaw) of virtual camera in the 3D space. Since chairs only swivel horizontally, the direction of vector is usually limited in X axis. However, the vertical optical flow is also measurable as far as the chair mechanically supports rocking motion, which can be naturally translated as the pitching movement of virtual camera.

4.7 Discussion

In contrast with earlier work, the Interactive Chair can provide the following benefits. First, it does not require users to relate the device manipulation by hand to the locomotive action consciously in the virtual space, because it directly leverages our proprioceptive sense of body and spatial conception. Second, a user can control the locomotive speed dynamically by leaning forward or backward with a different amount of degrees. This restores the natural correlation between our egomotion (the movement of the observer himself) and surroundings in the virtual environment, generating ambient optic arrays which otherwise can be experienced in the physical world only. Third, the use of torso for virtual locomotion frees both hands to manipulate other devices, so that more intellectual activities (e.g. typing, gesturing, grabbing or pointing) can be concurrently supported. Finally, the small amount of torso motion required by Interactive Chair does not quickly increase the fatigue of users, compared to prior 3D locomotion techniques that are based on full-body movement of the user.

Technically, the Interactive Chair currently allows 2DOF (forward/backward shifting and yawing), whereas standard game interfaces usually provide 3DOF for locomotion (forward/backward shifting, left/right shifting and yawing) and another 1DOF for viewing (pitching). We can achieve additional 1DOF (left/right shifting) in many different ways. For example, a grid of force sensors can be arranged on the seat plane, so that the center of gravity from the occupant's body can be measured. As the user moves his/her torso, the center of gravity will follow the same direction of movement. This design can provide 2DOF locomotion (forward/backward and left/right), however tracking information from this method can be easily biased by any swiveling movement of the chair that occurs concurrently. Meanwhile, the use of an IR sensor and reflective markers can be also considered as a method to obtain 2DOF locomotion. In this case, the user needs to attach three reflective markers on the back, so that the IR sensor on the backrest can measure both size and center position of the triangle formed by those markers. While the relative movement of the triangle within the sensor frame gives the information about left/right shifting, the size of triangle indicates the relative distance between the user and the IR sensor (i.e. forward/backward shifting).

Note that the system can also detect left-right leaning postures from another set of optical flow fields generated by the user's motion relative to the chair. This would be done with a secondary webcam on the backrest of chair, facing towards the back of the user. However, detecting optical flow as a way of motion estimation could result in potential drifting errors, since the magnitude of optical flow vector can vary depending on the webcam's FOV, the distance between camera and objects, etc. Alternatively, a gyroscope sensor or electric compass can be used for tracking swiveling motion. Haptic feedback is another element to improve the natural experience with virtual environments. Arranging and controlling a grid of vibrotactile devices on the chair can provide the user with location-specific feedback from the virtual space.

CHAPTER 5 MOTOR-INTENTIONAL MANIPULATION OF VIRTUAL OBJECTS

5.1 Grasping, Pointing and Motor Intentionality

We have discussed so far ways in which we visually experience the world (no matter whether it is physical or virtual), and how we naturally use our own body to succeed doing it. While such bodily acts fulfill our intrinsic needs to perceive and understand surroundings, we often exploit another kind of perception-action interplay so as to actively engage with the environment, for instance, touching and manipulating things in the world.

In fact, philosophers found that perceiving and acting upon an object in the world are more basic modes of intentionality rather than having thoughts about it. Merleau-Ponty describes that every event related to movement or sense of touch causes consciousness to put up a host of intentions, which run from the body as the center of potential action either towards the body itself or towards the object. In order to convert the thought of a movement into an actual movement, according to him, something exists between movement as a third person process and thought as a representation of movement – something which is an anticipation of, or arrival at, the objective and is ensured by the body itself as a motor power, a ‘motor intentionality.’¹ From the case of a patient with visual pathology, he found that the actions of grasping and touching are canonical motor-intentional activities. Given that this patient was unable to point to some part of his body unless he is allowed to take hold of it, Merleau-Ponty construed:

*(...) If the patient is set the task of interrupting the movement before its completion, or if he is allowed to touch his nose only with a wooden ruler, the action becomes impossible. It must therefore be concluded that ‘grasping’ or ‘touching’, even for the body, is different from ‘pointing.’ From the outset the grasping movement is magically at its completion; it can begin only by anticipating its end, since to disallow taking hold is sufficient to inhibit the action.*²

More recently, Kelly paid special attention to this difference between grasping and pointing observed by Merleau-Ponty, and reemphasized that our skillful, unreflective bodily activities, such as grasping the doorknob (in order to go through the door), are not the same as our reflective, cognitive or intellectual acts, such as pointing at the doorknob (in order to identify it).³ Kelly defines this motor intentionality as a behavioral phenomenon between the mechanical and the cognitive, by denying the empiricism in which mere reflex movements cannot be distinguished from directed skillful motor actions, as well as by rejecting the cognitivism in that unreflective motor actions (e.g., grasping an object) cannot be differentiated from deliberate, cognitive actions (e.g., pointing at an object).⁴

Such skillful and unreflective activities involve our bodily, situational understanding of space and spatial features. In the domain of neuroscience, researchers demonstrated that there are two different streams of visual information flow in the brain, one of which is geared to

¹ Merleau-Ponty, M. (2003 [1945]). *Phenomenology of Perception*. Routledge Classics, 127.

² *Ibid.*, 118.

³ Kelly, S. D. (2004). *The logic of motor intentional activity*, Draft.

⁴ Kelly, S. D. (2000). Grasping at Straws: Motor Intentionality and the Cognitive Science of Skillful Action. In *Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus (II)*. Eds. Mark Wrathall and Jeff Malpas. MIT Press, Cambridge, MA, 161-177.

perceptual judgment, the other of which is geared directly to action. Kelly refers to Milner and Goodale's work, among others, as it opened the way for acceptance of Merleau-Ponty's distinction between bodily understanding of spatial features and cognitive understanding of them. In their observation of a patient with visual pathology, they found the patient is poor at perceptual report of object qualities (e.g., the orientation of a rectangular slot on a disc), but still is good at using those same qualities to guide her actions (e.g., moving her hand towards the slot in the correct orientation and inserting it accurately).⁵ They attributed this result to the dissociation between perceptual-report tasks and visuo-motor ones, suggesting that there is no single, common understanding of space based on which both judgment and action occur, but rather two different ways of understanding of spatial qualities. From their work, Kelly sees that this patient has a motor intentional understanding of orientation, in Merleau-Ponty's term, since she understood the orientation based on her bodily capacities and dispositions to act with respect to it.⁶

Recent psychological studies are in agreement with this view. Paulignan & Jeannerod⁷ identified the notion of separate visuo-motor channels, which process visual information as required by either the transport or manipulation component of goal directed hand movements. Bekkering and Neggers⁸ demonstrated that intended grasping tasks improve the behavior of selecting object orientation better than pointing tasks do. Fischer and Hoellen⁹ investigated the impact of different motor demands on space- and object-based attention allocation, and found more object-based attention is required for grasping than for pointing. Their results are in line with brain image studies, which suggest that there are dedicated pathways from perception to action for grasping and pointing. Faillenot and colleagues demonstrated, in the grasping condition, a selective cerebral blood flow increase in the anterior part of the inferior parietal cortex and part of the posterior parietal cortex.¹⁰ Jeannerod and colleagues also found "grasping" neurons in the inferior parietal lobule and the inferior premotor area, which code size, shape, and orientation of objects and the specific grip types needed to grasp them.¹¹

5.2 Interaction Techniques for Virtual Object Manipulation

5.2.1 Pointing as an object selection technique

We have discussed that our everyday interaction with physical objects around us is pretty much motor-intentional. In the current practices of virtual environments, however, such a skillful mode of interaction is not sufficiently supported. Particularly, the manipulation of 3D objects is primarily based on 'pointing' techniques, the motivation behind which is to allow the user to

⁵ Milner, A. D. and Goodale M. A. (2006) *The Visual Brain in Action*. Oxford University Press, NY. 126-128.

⁶ Kelly, Op.cit.

⁷ Paulignan, Y. and Jeannerod, M. (1996). Prehension movement: The visuomotor channels hypothesis revisited. In Wing, A. M., Haggrad, P. and Flanagan J. R. (eds.), *Hand and Brain*. Academic Press, NY. 265-286.

⁸ Bekkering, H. and Neggers, F. W. (2002). Visual search is modulated by action intentions. In *Psychological Science* (13), 370-374.

⁹ Fischer, M. H. and Hoellen, N. (2004). Space- and Object- Based Attention Depend on Motor Intention. In *The Journal of General Psychology*. 131(4): 365-377.

¹⁰ Faillenot, I., Toni, I., Decety, J., Gregoire, M. C., and Jeannerod, M. (1997). Visual pathways for object-oriented action and object recognition: Functional anatomy with PET. In *Cerebral Cortex* (7), 77-85.

¹¹ Jeannerod, M., Arbib, M. A., Rizzolatti, G., and Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. In *Trends in Neurosciences*. 18(7): 314-320.

conveniently select and manipulate objects located beyond the area of reach by pointing at them.¹² Poupyrev and colleagues¹³ implemented this idea by developing so-called virtual ray-casting technique (Figure 5.1), with which the user points at objects with a virtual ray that defines the direction of pointing. In some cases, this virtual ray is replaced with a conic volume to provide a softer selection method that does not require high precision and accuracy of pointing. Flashlight¹⁴ technique, for instance, is implementing this approach.

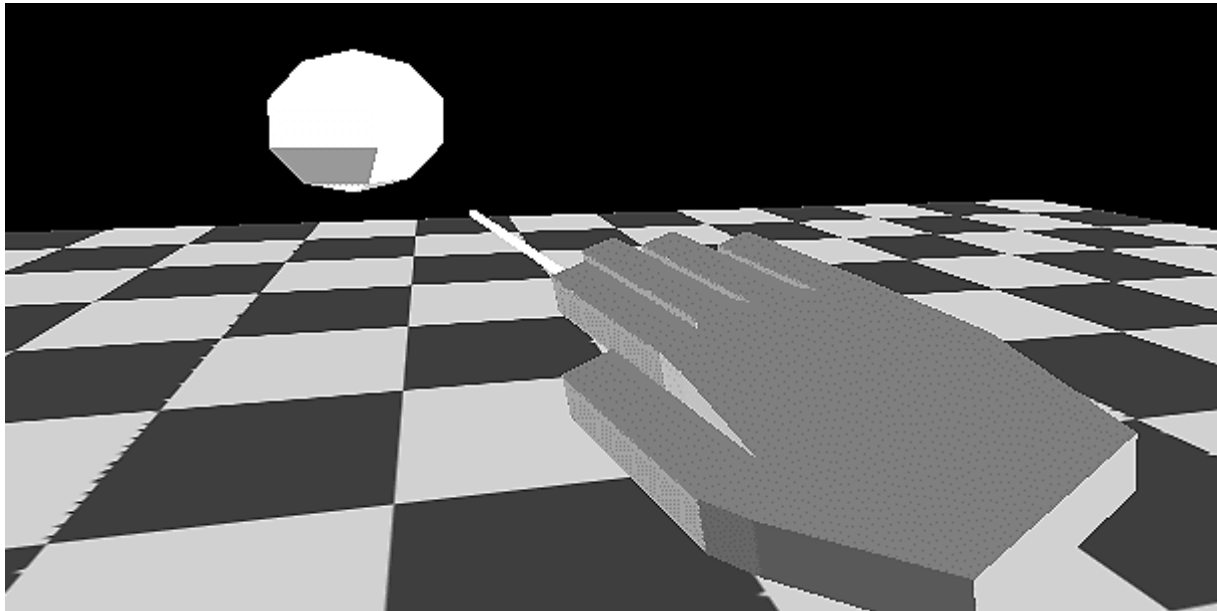


Figure 5.1: Virtual ray-casting technique

In general, pointing has been considered as a powerful selection technique in the domain of HCI, since it requires much less physical (hand and/or arm) movement from the user, than direct manipulation techniques.¹⁵ A mouse interface is the most common pointing device these days, and is often used, not just for 2D programs, but also for 3D applications. There also exist other types of pointing devices, including digital pen (stylus), inertial and optical tracking devices, etc.

However, this pointing-based approach relies on our cognitive and intellectual acts, in phenomenological terms, ignoring the motor-intentional property of our daily activities in the physical world. In this type of interaction mode, manipulation of a virtual object normally takes multiple steps of mediation or interpretation for the user to complete. For example, moving an object with an input device consists of a set of cognitive mapping between 1) the device's position in the space (or on the surface, in case of 2D input device) and the cursor position on the screen, 2) the button-clicking action and the virtual motion of grabbing the object, 3) the motion

¹² Bowman, D. A., Kruijff, E., LaViola Jr., J. J., and Poupyrev, I. (2005). *3D User Interfaces: Theory and Practice*. Addison Wesley.

¹³ Poupyrev, I., Weghorst, S., Billingham, M. and Ichikawa, T. (1998). Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. In *Computer Graphics Forum, EUROGRAPHICS Issue 17(3)*: 41-52.

¹⁴ Liang, J. and Green, M. (1994). JDCAD: A Highly Interactive 3D Modeling System. In *Computers and Graphics*, 18(4): 499-506.

¹⁵ Poupyrev et al., Op.cit.

of moving the device and the movement of the virtual object, and finally 4) the button-releasing event and the motion of the object being dropped. This complex process keeps our engagement and involvement with the virtual content somewhat limited, and gives us a constant feeling of remoteness during the manipulation of virtual objects. In this aspect, the aperture selection technique¹⁶ developed by Forsberg and colleagues is worth to pay attention to, because their approach simplifies such cognitive process by aligning both the pointing device and the virtual object with the user's line of sight. As shown in Figure 5.2, the apex of the conic volume is set to the location of the participant's dominant eye, and the direction vector of the cone is the vector from that eye through the tracker's location (represented by a cursor). However, since this technique is still based on the 'pointing' metaphor, users were allowed to manipulate virtual objects only with a drumstick prop, and it was assumed that the user's eye position is relatively fixed to the position of display device.

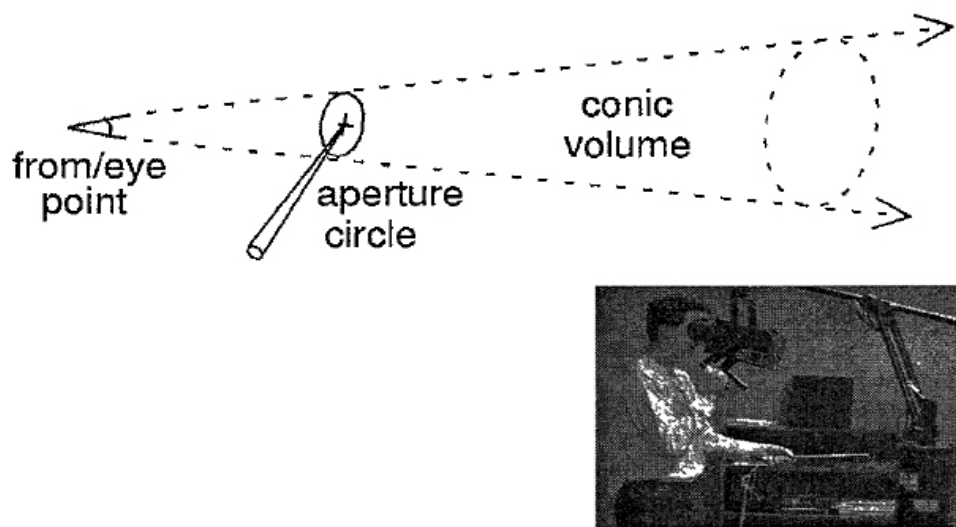


Figure 5.2: Aperture selection technique

5.2.2 Virtual hand techniques

While the aforementioned pointing techniques control the cursor on the screen in 2D space only, the virtual hand techniques map the position and orientation of the input device onto the position and orientation of a 3D cursor, which is usually represented as a 3D model of a human hand. The input device can be data gloves, 3D mice, or other types of tracking devices. To select an object, the user intersects the 3D cursor with the target object, and then uses other trigger methods (e.g., button, gesture, etc.) to pick it up. There have been many different variations of this approach, however in most cases the 3D cursor and the input device are not visually aligned with the user's line of sight (unless the user wears a head-mounted display device), too. Even in the most recent developments, such as Oblong Industries' spatial operating environment named g-speak (Figure 5.3), the coordination between the hand, the virtual object, and the user's viewpoint has not been

¹⁶ Forsberg, A., Herndon, K., and Zeleznik, R. (1996). Aperture Based Selection for Immersive Virtual Environments. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, 95-96.

seriously considered.¹⁷ Such a spatial misalignment between what the user sees and what the user does makes our 3D interaction modalities more dependent to pointing techniques, therefore the motor-intentional nature of physical object manipulation is not taken into the consideration.



Figure 5.3: Oblong Industries' G-Speak technology

5.3 Direct and Motor-Intentional Manipulation of Virtual Objects

In order to avoid the cognitive mapping process induced by pointing techniques and utilize our motor-intentional understanding of object and space, I propose a method that provides the sense of 'grasping' rather than 'pointing.' In this approach, the manipulation of a virtual object can happen when the user's hand and the object are on the user's line of sight, as depicted in Figure 5.4. From the user's viewpoint, the hand and the virtual object are visually aligned in space, and accordingly it affords the sense of the virtual object being grasped by the user's hand. To obtain this alignment, the system needs to track both the user's hand and his face position in 3D space, and calculate the line of sight from the user's viewpoint. In the prototype implementation, I track the user's fingers with an IR sensor-emitter and a reflective marker set, and determine three types of object manipulation by hand: grasping, moving and releasing an object.

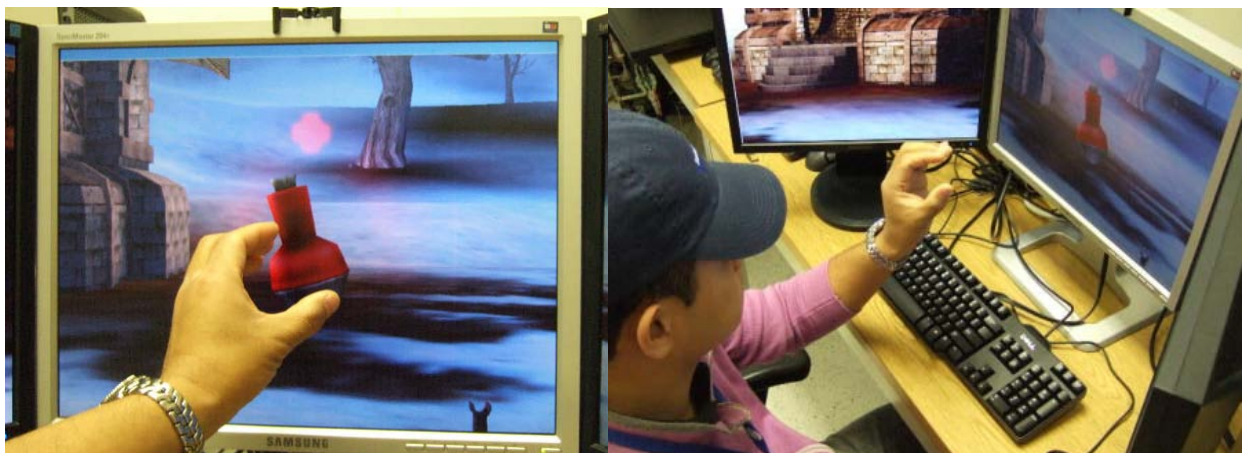


Figure 5.4: In-line mediation of object manipulation

¹⁷ Oblong Industries. WWW URL: <http://www.oblong.com>

5.3.1 Finger Tracking and Hand Pose Detection

In contrast to existing hand gesture tracking devices, I only track the user's thumb, forefinger, and middle finger at their tips. This is because 1) they are relatively important in the action of true grasping, and 2) with the hand in a relaxed (neutral) position, their tips tend to form a triangular shape, which helps disambiguating complex hand poses that can be observed during the direct manipulation of an object. In order to avoid any intrusive devices to be worn by the user, I use a minimal set of reflective markers and an IR sensor-emitter for tracking fingers. The proposed system detects the position of three markers attached on fingertips by using the IR sensor, and infers the position (X, Y) of the hand on the virtual image plane by calculating the centroid of the triangle formed from those three dots (markers). Figure 5.5 demonstrates how this approach works.

In the mean time, I use the size (area) of the triangle to determine the state of grasping or releasing. When the area of triangle decreases below the predefined threshold Gr , it is assumed that the hand gets into the grasping condition. The value of threshold Gr should be, in principle, represented as:

$$Gr = C / d^2$$

, where C is a constant, and d is the distance between the image plane of the IR sensor and the centroid of finger tips. However, for rapid prototyping of the system, I assumed that the user's hand is placed approximately halfway between the screen (where the IR sensor is mounted) and the user's face, to simplify the calibration process.

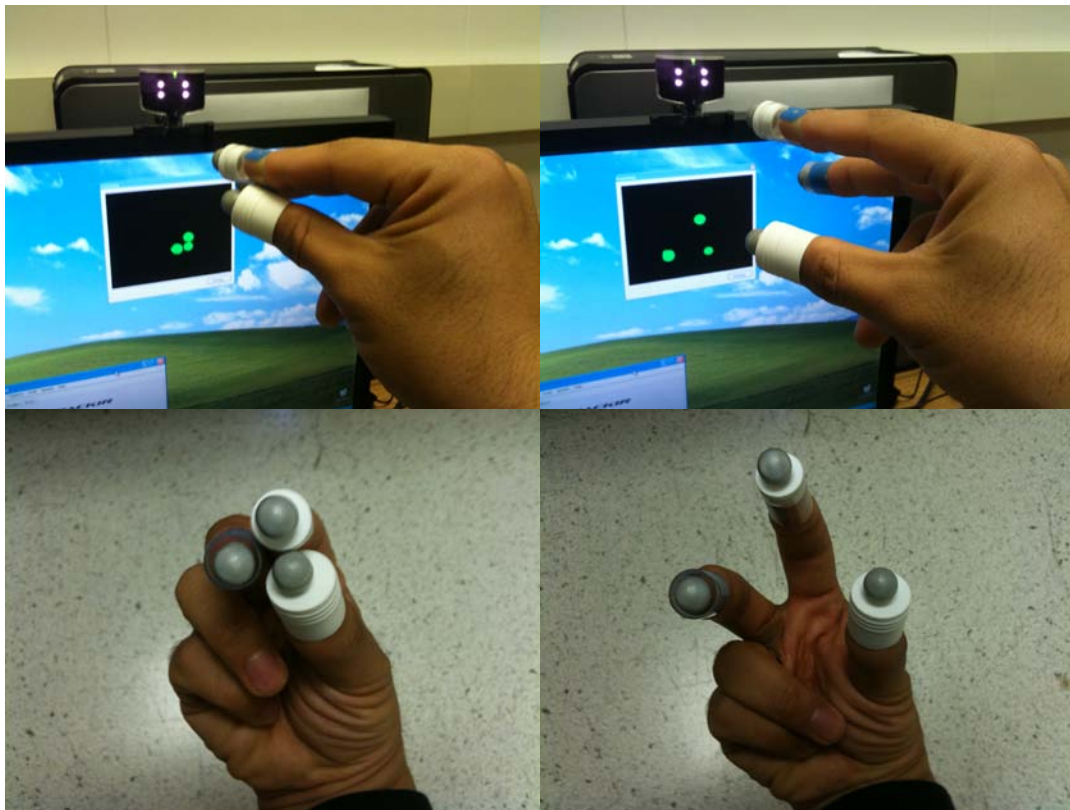


Figure 5.5: Finger tracking and hand poses
(left column: grasping hand pose / right column: relaxed (released) hand pose)

5.3.2 Line-of-Sight Coordination between a Hand and a Virtual Object

Once the user's fingers (as well as face) are tracked, the proposed system transforms both 2D face tracking space and 2D hand tracking space into 3D model space. As we discussed in the earlier chapter, the face position (x_f, y_f) in the size of r is translated into the 3D coordinates:

$$(c_1(x_f - 160), c_1(120 - y_f), c_2(r_d - r))$$

, where c_1 and c_2 are calibration factors and r_d is the default size of the face. In the same manner, the position of centroid of fingers (x_c, y_c) is translated into the 3D coordinates:

$$(c_3(x_c - 175), c_3(144 - y_c) + c_4, c_2(r_d - r) - n_c/2)$$

, where c_3 and c_4 are calibration factors and n_c is the distance from the virtual camera to the near clipping plane. The resolution of IR sensor plane is 350 X 288 pixels.

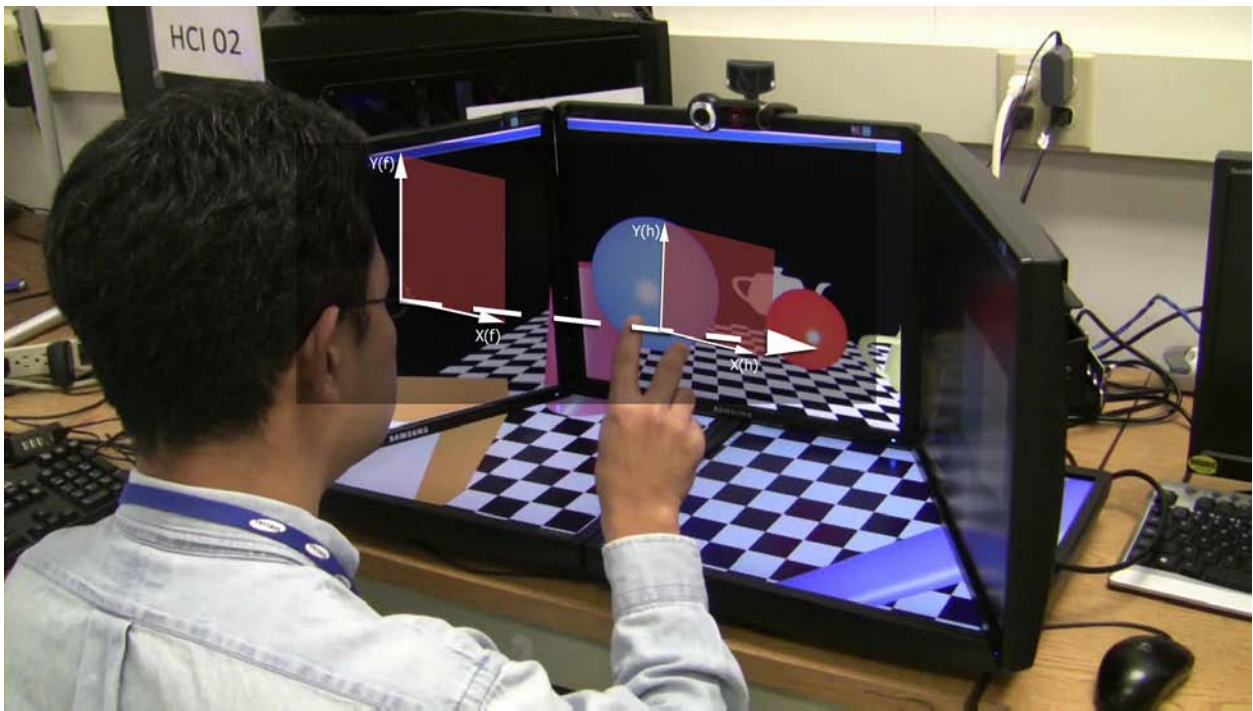


Figure 5.6: Line-of-sight coordination between the user's hand and a virtual object

As shown in Figure 5.6, the system then determines which object on the screen is aligned with the user's face and fingers, by casting a virtual ray from the center position of the user's face (assumed as the eye level), through the centroid of the fingers, onto the screen surface. When this ray intersects with one of virtual objects while the fingers are in the grasping condition, that object is picked up and follows the movement of tracked fingers until the fingers return back to the releasing state. Since this method tracks both face and fingers, it can be successfully integrated with the view-dependent rendering technique in the non-planar display space.

5.4 Discussion

The proposed interaction method promotes the motor-intentional nature of object manipulation behaviors such as grasping, moving and releasing an object, by coordinating our visual perception with our motor control. While I expect this mode of interaction to increase the degree of engagement with virtual content, it still has room for further improvement. First, our visual perception is accurate enough to notice the difference in depth, between the hand and the virtual object; therefore to certain extent the cognitive adjustment is still necessary to be able to interact with the virtual content. We could consider the use of stereoscopic rendering technologies to overcome this limitation, by further coordinating the binocular disparity factor with the positions of face, hand and virtual objects in 3D space. Second, the motor-intentional act of grasping normally comes with, or is followed by, the sense of touch, providing the sensory confirmation for grasping. Reconstructing this tactile feedback loop for the user with bare hand is technically challenging, but this multimodal approach will significantly enhance the sense of involvement with the virtual environment.

CHAPTER 6 USER TEST AND EVALUATION

6.1 User Test Overview

The overall objective of the user test is to investigate if the interaction methods I proposed in the earlier chapters can improve our experience with virtual environments. In this experiment, I assume that three factors related to natural perception-action loop, individually and/or in combination, contribute to the quality of interaction with virtual worlds. The first factor I focus on is the correlation between the user's perspective and (content) representation. My expectation is that the view-dependent rendering of 3D content on the non-planar display setup enhances the user's field of view as well as the visual depth perception (motion parallax) simultaneously. The second factor is the coordination between the sense of body and locomotion. I assume that the perception of self-body movement during the locomotion control is beneficial to our understanding of virtual space, as this approach alleviates the complex cognitive mapping between the manual interfaces (mouse, keyboard, gamepad, etc.) and the virtual navigation. The third is the alignment of hand motions with content on the screen. I expect that the in-line mediation of hand poses into the virtual object manipulation is more natural than the indirect, remote control of virtual objects using a mouse, since this in-line mediation supports our motor-intentional behaviors which have been ignored in the current practice of 3D interface design. In order to evaluate the validity of these factors independently, a separate user test was performed for each factor, with each pertinent prototype system developed earlier in this dissertation work.

I adopted the experiment procedures from well-established information technology literatures. Particularly, the first and second experiments mainly refer to three categories of navigation tasks defined by Bowman and colleagues¹, including exploration, search and maneuvering tasks. In an *exploration task* the user is investigating the surroundings with no special target in mind. In a *search task* the user is moving to reach a special target location. A *maneuvering task* is performed when the user wants to give the viewpoint a more advantageous position and orientation for carrying out a specific task. I am going to examine the subjects' performance quantitatively during the first two experiments in these aspects.

In addition, at the end of each experiment, I also measure the qualitative user experience by asking subjects to fill out the questionnaire form shown in Table 6.1. I carefully selected 9 items of questionnaire from Witmer and Singer's work on measuring presence², and sorted them into three categories of user experience factors – naturalness, involvement and mastery. I assume that each of these categories is closely related to the sustainability of perception-action loop in the virtual environments. In every session of the experiments, I aggregate the subject's feedback to each questionnaire item to create a session response, by using a 7-point Likert scale (e.g., 1= not well, 4=moderately well, 7=very well).

¹ Bowman, D. A., Kruijff, E., LaViola, J. J., Jr., and Poupyrev, I. (2001). An introduction to 3-D user interface design. In *Presence: Teleoperators and Virtual Environments*, 10(1): 96–108.

² Witmer, B. G. and Singer M. J. (1999). Measuring Presence in Virtual Environments: A Presence Questionnaire. In *Presence: Teleoperators and Virtual Environments*, 7(3): 225-240.

Factor Category	ID	Questionnaire Item	
Naturalness	1	How natural did your interactions with the environment seem?	
	2	How much did your experiences in the virtual environment seem consistent with your real-world experiences?	
	3	.1	How well could you examine objects from multiple viewpoints?
		.2	How well could you move around inside the virtual environment?
		.3	How well could you move or manipulate objects in the virtual environment?
Involvement	4	How involved were you in the virtual environment experience?	
	5	Were you involved in the experimental task to the extent that you lost track of time?	
	6	Were you able to anticipate what would happen next in response to the actions that you performed?	
Mastery	7	How much were you able to control events?	
	8	How quickly did you adjust to the virtual environment experience?	
	9	How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	

Table 6.1: Post-questionnaire items in three categories

As the experimental setup, I prepared three independent prototype systems, as shown in Figure 6.1. In the first experiment, each subject is seated before the non-planar display consisting of 3 vertical LCD monitors (20”), with the webcam attached for face tracking. In the second experiment, I provide only one LCD monitor and the Interactive Chair. In the third experiment, I use a normal desktop environment except the mouse removed, and provide an IR sensor-emitter and reflective marker sets instead.



Figure 6.1: Experimental systems (from left: for experiment 1, 2 and 3)

On the other hand, as the control setup, I use a typical desktop computer setting with a mouse and a keyboard, as the control environment. For the first and the second experiments, it provides motion controls in 4 dimensions (DOF) listed in Table 6.2, which are typically used in most FPS (First Person Shooter)-style 3D games or virtual world applications.

Manual Control	Action in 3D Space	Dimension
'S' or '↑' key	Move forward	Z
'S' or '↓' key	Move backward	
'A' or '←' key	Move left	X
'D' or '→' key	Move right	
Horizontal (Left-Right) Mouse Movement	Turning left-right	YAW
Vertical (Forward-Backward) Mouse Movement	Turning up-down	PITCH

Table 6.2: Mouse mapping in the control system

In the third experiment, the subjects in the control group only use a mouse, with which they can pick a virtual object on the screen and move it around in X and Z dimensions by dragging it with the mouse.

6.2 Experiment 1: View-Dependent 3D Content in Non-Planar Display Space

The goal of the first user study was to examine the effect of direct perceptual acts on the user's spatial perception in the virtual environments. I hypothesized that the view-dependent, non-planar display design improves the user's exploration task by enhancing the field of view and the relative depth perception. Eight individuals (3 females and 5 males) were recruited as subjects for this experiment, and each of them participated in both experimental and control groups in turns. Four of them tried out the experimental system first, and then repeated the same test for

the control system with later. The other 4 subjects went through the test in reverse order with the opposite content.

6.2.1 Test Procedure

Each session of the experiment was conducted in the following procedure:

- 1) The investigator gives the subject a brief overview of the experiment as well as instructions about how to use the given system.
- 2) The subject freely plays with the system for 5 minutes, so as to get used with the user interface of the system.
- 3) The investigator informs the subject that there will be five objects (sphere, cylinder, cone, box and teapot) hidden behind virtual walls, and ask them to identify each of them.
- 4) The session begins by pressing the spacebar on the keyboard.
- 5) A set of five virtual walls begin to move from the off-screen area through peripheral zone to the central area of the display space. These walls are eventually deployed at five different positions (front, top, bottom, left and right).
- 6) The subject starts exploring the 3D space and verbally reports to the investigator each time s/he identifies an object behind each virtual wall.
- 7) The session is over as soon as the subject reports the last (fifth) object identified.
- 8) The subject fills out the post-questionnaire.

I have decided to let the subjects verbally report identified objects, because it is a direct, immediate and intuitive communication method that can minimize any potential delay during the reporting activity. In practice, subjects simply said “teapot”, for example, while pointing left with a finger simultaneously, meaning that there is a teapot behind the virtual wall on the left.

6.2.2 Performance Evaluation

The test results demonstrate that the experimental group was able to explore the given virtual space in a shorter amount of time, than the control group did. Figure 6.2 shows that the performance of each subject in both experimental and control environments. In average, subjects in the experimental groups completed the exploration task in about 29.1 seconds, while it took about 35.9 seconds for the control group. The standard deviation for the control group (12.8) was 5.0 higher than the experimental group’s one (7.8), meaning that the performance gap between skilled and casual users was higher in the control environment (I determined those who completed the control test within 30 seconds as skilled users). However, it turned out the experimental setup did not really improve the performance of skilled users (subject ID 1, 4 and 6). Therefore, I concluded that the non-planar display system with the view-dependent rendering method is more beneficial to non-skilled, casual users than skilled ones.

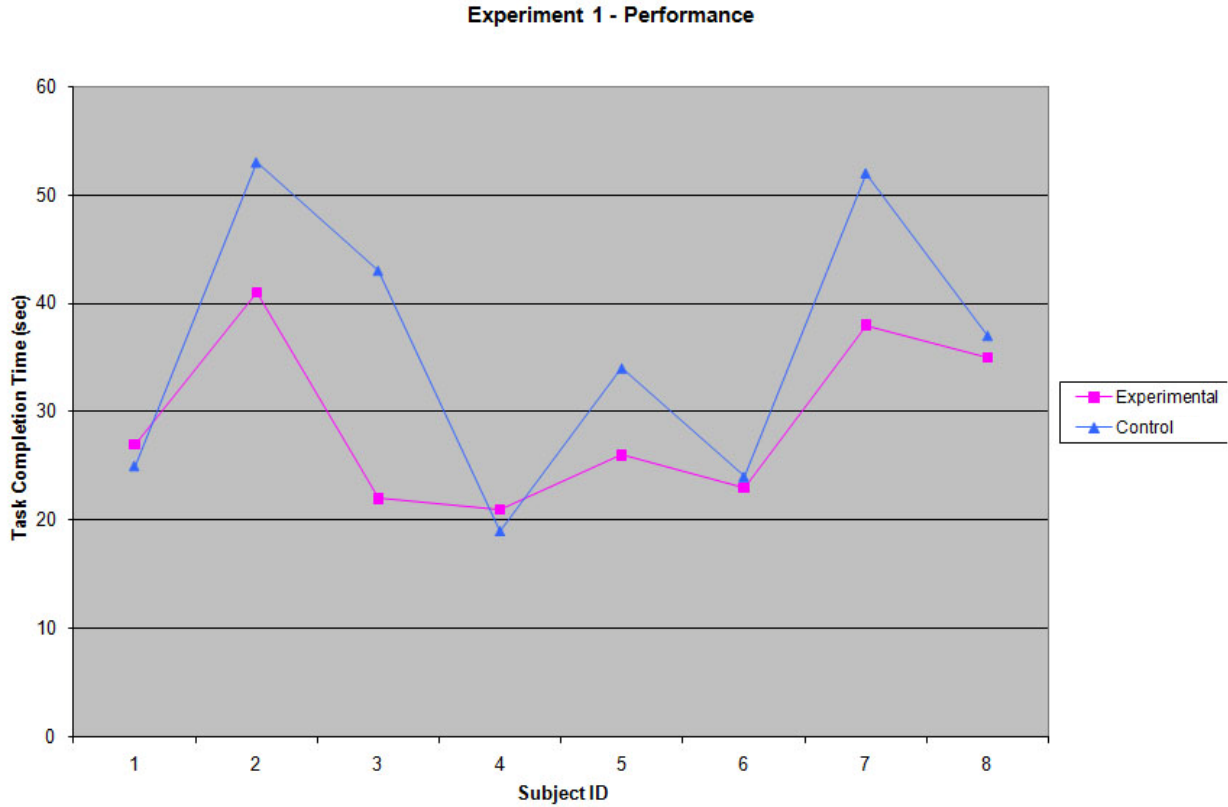


Figure 6.2: Experiment 1 performance comparison

6.2.3 Questionnaire Evaluation

Per each category of user experience factors, I compared the sum of Likert scores between experimental and control groups. Paired T-test was performed to determine the statistical significance of each comparison. The analysis showed that the experimental system was statistically more natural to use, than the control system ($t(8) = 3.05, p < .01$). The experimental system also delivered a better sense of involvement to subjects ($t(8) = 1.72, p < .1$). However, it showed that the feeling of mastery was weaker when using the experimental system ($t(8) = -2.73, p < .05$) (figure 6.3).

Experiment 1 - Questionnaire

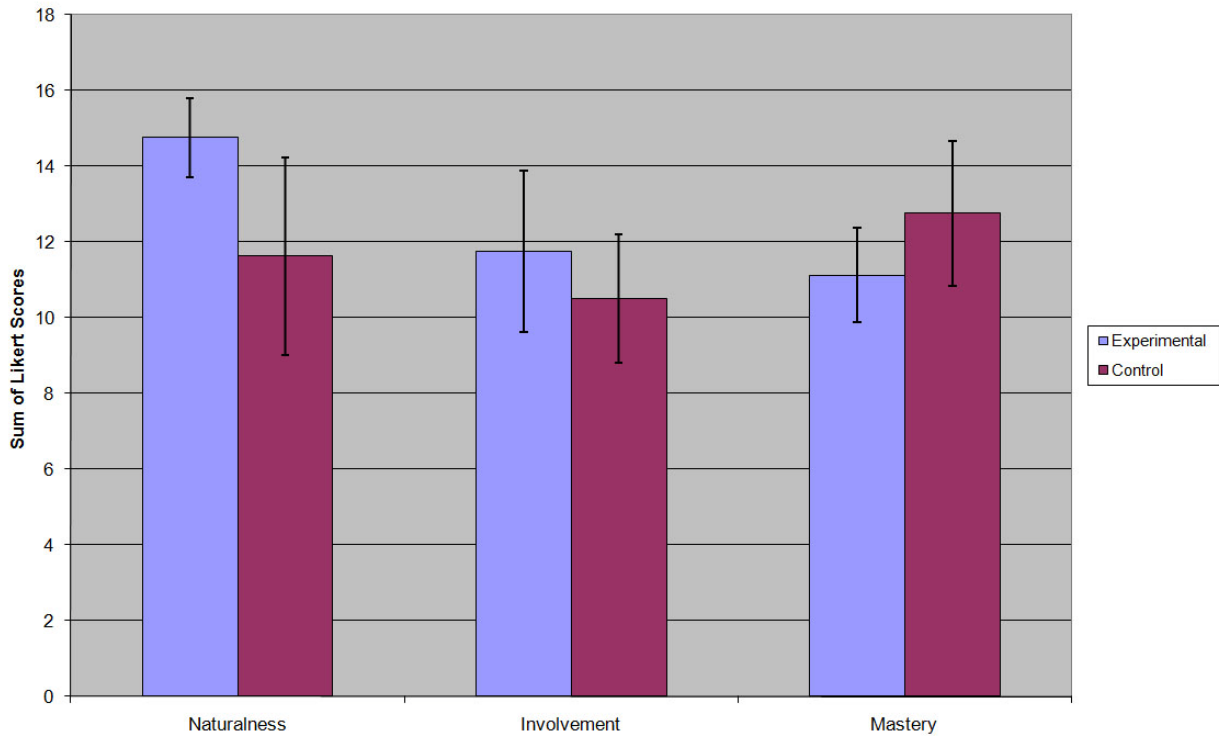


Figure 6.3: Experiment 1 questionnaire (T-test)

6.3 Experiment 2: Proprioceptive Locomotion Control with Interactive Chair

In the second user study, I examined the effect of proprioceptive sense on the locomotion in the virtual environment. I hypothesized that the proposed Interactive Chair system improves the user's search and maneuvering tasks by utilizing our subtle torso movements, which are being captured in the sitting position but are also normally associated with our walking behaviors. Eight individuals (3 females and 5 males) were recruited as subjects for this experiment, and each of them participated in both experimental and control groups in turns. Four of them tried out the experimental system with the clockwise travel route first, and then repeated the same test in the control setup, with the counterclockwise route later. The other 4 subjects went through the test in reverse order with the opposite direction of travel route.

6.3.1 Test Procedure

Each session of the experiment was conducted in the following procedure:

- 1) The investigator gives the subject a brief overview of the experiment as well as instructions about how to use the assigned system.
- 2) The subject freely plays with the system for 5 minutes, so as to get used with the user interface of the system.
- 3) The subject is instructed to travel along a planned route by using the given system interface. The route included 3 towers, 3 houses and 1 rock on its passage as shown in Figure 6.4.

4) The session begins when the subject's avatar departs from the first building (the starting point).

5) The session is over as soon as the avatar arrives at the last building (the final destination).

6) The subject fills out the post-questionnaire.



Figure 6.4: Route for virtual traveling

The travel route was designed as simple as possible, in order to alleviate the subject's effort to find the way to the destination. According to Bowman and colleagues³, navigation in a virtual environment can have two meanings; a motor aspect called *travel* and a cognitive aspect called *wayfinding*. While travel is the movement of the viewpoint from one location to another, wayfinding can be described as the cognitive process of determining a path through the environment to the desired destination. Since my focus is on investigating the coordination between proprioceptive motor control and locomotion, the complexity of travel route had to be minimized.

³ Bowman, et al., Op.cit.

6.3.2 Performance Evaluation

The test results demonstrate that the experimental group was able to travel the virtual space along the given route in a shorter amount of time, than the control group did. Figure 6.5 shows that the performance of each subject in both experimental and control environments. In average, subjects in the experimental groups completed the search and maneuvering task in about 80 seconds, while it took about 95 seconds for the control group. The standard deviation for the control group (10.27) was 2.31 higher than the experimental group's one (7.96). Similarly to the first test, the experimental setup improved the performance of casual users better than the performance of skilled users (subject ID 3, 4 and 5, who completed the control task within 85 seconds).

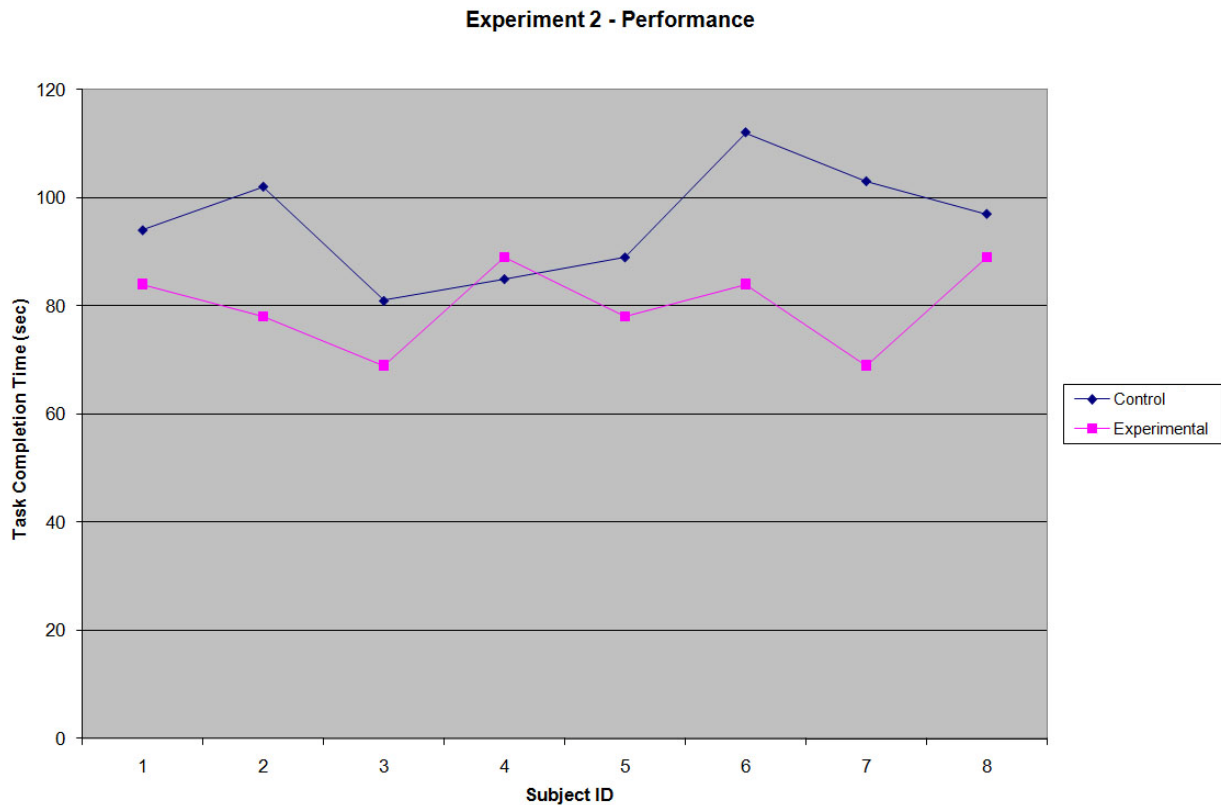


Figure 6.5: Experiment 2 performance comparison

6.3.3 Questionnaire Evaluation

From the subjects' responses to the post-questionnaire, I ran the paired T-test to determine the statistical significance of comparison between experimental and control groups. The analysis showed that the experimental system was statistically more natural to use, than the control system ($t(8) = 0.74, p < .25$). The experimental system also delivered a better sense of involvement to subjects ($t(8) = 0.87, p < .21$). Unlike the first experiment, the feeling of mastery was stronger when using the experimental system ($t(8) = 1.64, p < .1$) (figure 6.6).

Experiment 2 - Questionnaire

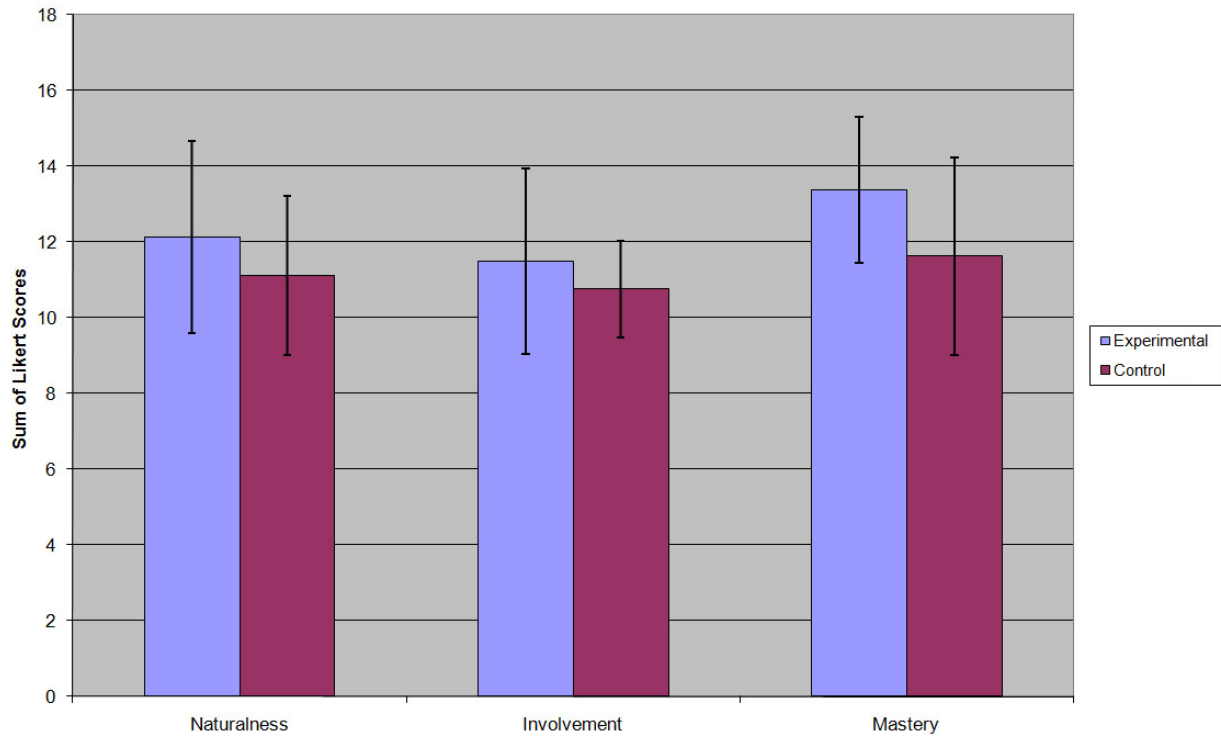


Figure 6.6: Experiment 2 questionnaire (T-test)

6.4 Experiment 3: Motor-Intentional, In-line Manipulation of Virtual Objects

The third user study was aimed at examining the effect of motor-intentional behaviors on the interaction with virtual content. I hypothesized that, when it comes to the manipulation (grasping, moving and releasing) of a virtual object, the direct hand poses over a visually-aligned virtual object is more natural and intuitive than the indirect control using a mouse, since it induces our motor intentionality rather than cognitive interpretation. Six individuals (2 females and 4 males) were recruited as subjects for this experiment, and each of them participated in both experimental and control groups in turns. Three of them tried out the experimental system first, and then repeated the same test for the control system later. The other 3 subjects went through the test in reverse order.

6.4.1 Test Procedure

Each session of the experiment was conducted in the following procedure:

- 1) The investigator gives the subject a brief overview of the experiment as well as instructions about how to use the assigned system.
- 2) The subject freely plays with the system for 5 minutes, so as to get used with the user interface of the system.

3) The subject is instructed to grasp each of 4 objects from its original location, and move it to the opposite side of the display space (i.e., moving 2 objects from left to right side, and another 2 objects from right to left side).

4) The session begins as the subject grasps the first object.

5) The session is over as soon as the subject drops off the last (fourth) object at destination.

6) The subject fills out the post-questionnaire.

6.4.2 Performance Evaluation

The test results demonstrate that the performance of subjects in the experimental group was lower than the performance of subjects in the control group. Figure 6.7 shows that the performance of each subject in both experimental and control environments. On average, subjects in the experimental groups completed the manipulation task in about 52.3 seconds, while it took only about 29.3 seconds for the control group. The standard deviation for the experimental group was 7.77, and 8.52 for the control group. No obvious relation was found between the amount of performance improvement and the subjects' skill levels.

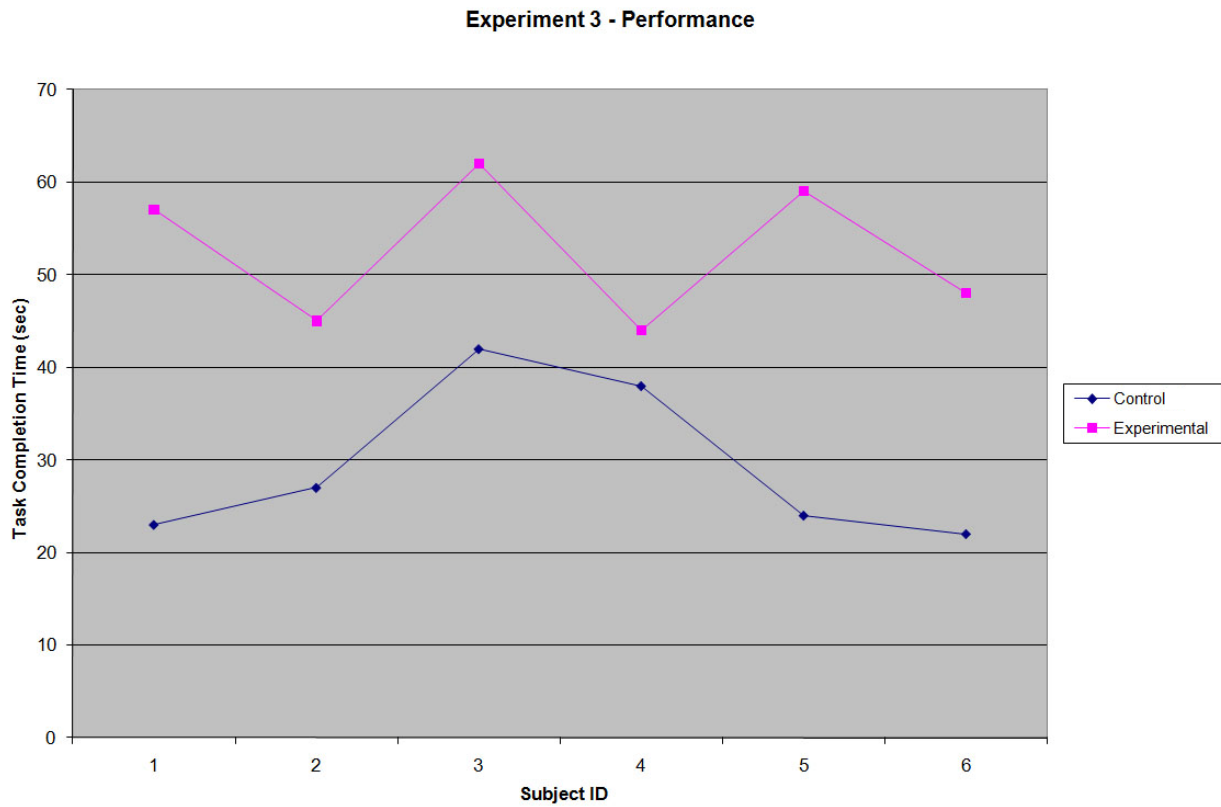


Figure 6.7: Experiment 3 performance comparison

6.4.3 Questionnaire Evaluation

I ran the paired T-test again from the post-questionnaire to determine the statistical significance of comparison between experimental and control groups' responses. The analysis showed that, although the performance of the experimental group was significantly lower than the performance of the control group, the experimental system was nearly as natural as the control system ($t(6) = -0.25, p < .5$), and also delivered a better sense of involvement to subjects ($t(6) = 1.08, p < .2$). However, in line with the performance test, the result showed that the feeling of mastery was significantly weaker, when using the experimental system ($t(6) = -5.21, p < .01$) (figure 6.8).

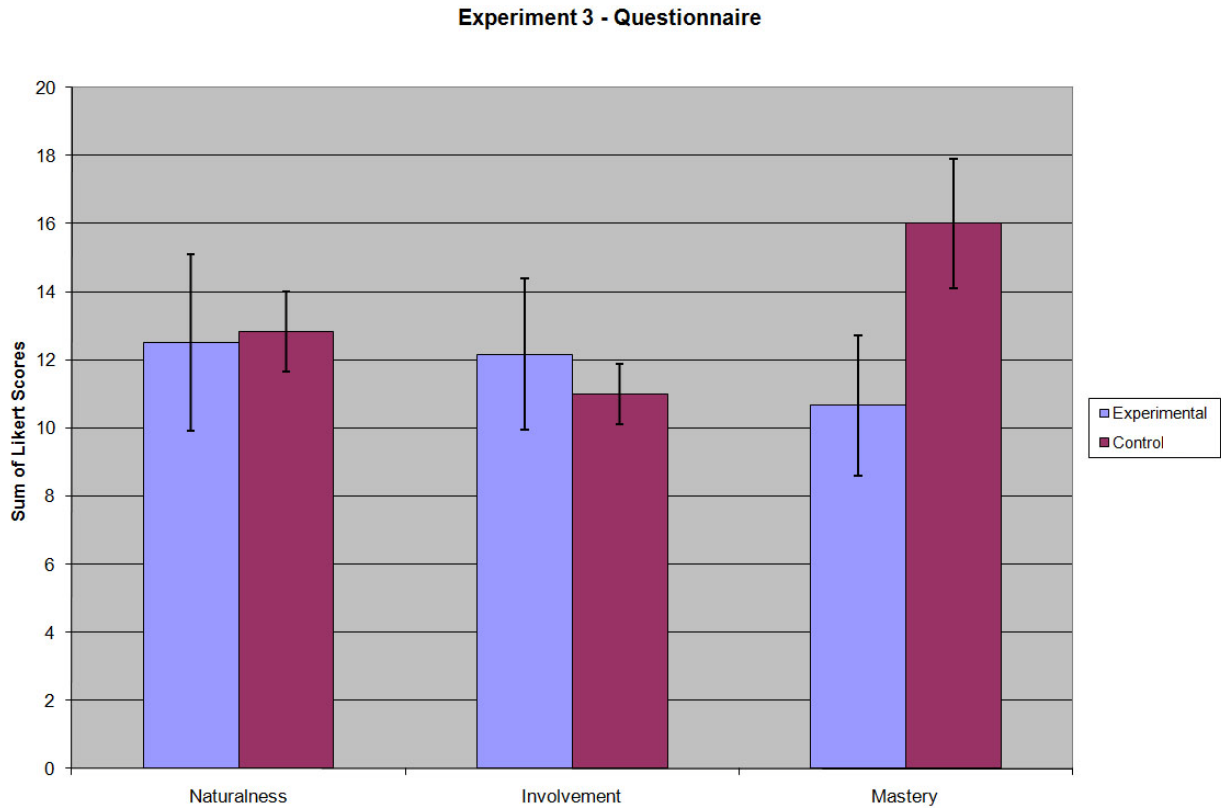


Figure 6.8: Experiment 3 questionnaire (T-test)

6.5 Discussion

The results of user test and evaluation demonstrate that the proposed interaction methods provide users with, in most cases, more natural and immersive experience of virtual environments. In the first two experiments, it turned out the user's bodily movements, which originated from our daily activities in the physical environment, improve the user's overall performance in the virtual world. Particularly, these methods were more efficient to the novice or casual users than the skilled users, as we have seen in the evaluation results. This means proposed interaction methods were easier to learn than the existing 3D interfaces. The mastery of the view-dependent, non-planar display system was below my expectation, since the face tracking module lost the

subject's face position intermittently and the current prototype only tracks the face in 3 dimensions (i.e., no yaw, pitch or roll).

The results show that the proposed object manipulation method decreased the subjects' performance significantly. I suppose that it is because 1) frequent hand movements in the air caused physical fatigue to the subjects more rapidly than the mouse manipulation on the desk surface, 2) the object grasping mechanism was not always as accurate as the pointing function of the mouse, due the limited precision in face and finger tracking, and 3) it was more difficult to grasp an object that is partially occluded by another one, than selecting it with a mouse. Nevertheless, subjects reported that the in-line object manipulation method was as natural as the mouse manipulation, and that they felt more involved in the virtual environment by grasping and moving an object directly.

CHAPTER 7 CONCLUSION

7.1 Conclusion

In this dissertation I have presented design and methods for natural interaction with virtual environments. I emphasized that the continuous perception-action loop is what keeps providing us with the sense of presence, and that this intrinsic loop is broken in the experience of virtual worlds. From a series of philosophical discussions on presence, perception and action, I have taken a phenomenological orientation as the basis for developing an approach to the design of natural and immersive interaction methods. In particular, I have focused on the following three embodiments of specific perception-action relations with which the virtual environments can extend our legitimate experience of being-in-the-world.

First, I paid special attention to direct perceptual acts, such as head turning and shifting, supporting the idea that our ability to perceive is constituted by our possession of bodily skill. I demonstrated how these acts can be mediated into the 3D virtual setting, by designing and implementing a non-planar display space and a non-intrusive view-dependent rendering technique. Technically, this approach enhances the user experience by extending the field of view and engendering the sense of depth, without requiring any large or head-mounted display. Moreover, from the user test, I demonstrated that this approach also increases the user's feeling of naturalness and involvement while interacting with a 3D virtual environment.

Second, I studied the locomotion as an extension of perceptual acts, following Gibson's position in which locomotion is treated as proprioceptive. I proposed a novel method to achieve natural navigation in virtual 3D space, by utilizing our natural sitting postures on a swiveling chair. The user test demonstrated this approach provides a more natural, immersive and easy-to-learn control of locomotion than the conventional mouse + keyboard combination does.

Finally, I considered the motor intentionality as a primary way to interact with objects in the world. By differentiating 'grasping' or 'touching' from 'pointing', I suggested the support of motor-intentional interaction with 3D content would improve the immersive sense of presence in virtual environments. I implemented and demonstrated a system that coordinates the object manipulation by hand with the user's line of sight, so that the user can feel like h/she is grasping the object in front of the screen. However the user test revealed the practical limitation of this approach, as the subjects' performance with this method was quite poor, even if they reported that it feels reasonably natural and involved.

Overall, from these three instances of perception-action loops, I demonstrated why philosophical and psychological insights need to be considered in studying natural and immersive experience of virtual environments, as well as how this approach help us identify limitations of current interaction methods and overcome them. I expect the result of this research to change how architects and HCI scientists will approach the design and development of virtual environments.

7.2 Future Research Directions

In the future, I plan to expand the scope of this research in the following aspects.

(1) Direct perceptual acts and binocular vision

While this dissertation has mainly focused on the action-based perception of space, we still cannot ignore the importance of binocular vision in our spatial perception. Stereoscopic display and rendering techniques have been advanced dramatically for the last decade, however most of

them do not work well, or even cause conflicts with, our natural perceptual acts, for instance, those for the motion parallax effect. As we discussed earlier, the natural and immersive sense of virtual presence does not only comprise of the content reality, but also the integrity of perception-action loop, therefore the successful support of both binocular and action-based perceptions needs to be seriously investigated.

(2) Object manipulation by motor intention

I have justified the necessity of motor-intentional interaction methods when it comes to the natural manipulation of virtual objects. In fact, the pointing metaphor has been dominantly applied to the design of human-computer interaction, thus forcing us to use only our cognitive faculties while suppressing the motor-intentional nature of interaction behaviors. In the mean time, realizing natural interaction with virtual objects by hand is a challenging task. While the precise tracking of hand and/or finger motions is an active research topic, providing necessary tactile feedbacks to the user's hand is also a non-trivial task. In addition, considering the stereoscopic display technologies in the design of object manipulation method is also an important research topic to be investigated.

(3) Interpersonal interaction in the virtual environments

Thus far we have only discussed about how we (as a first-person user) interact with space and objects that virtually surround us. Ultimately, those findings from this dissertation can be, and need to be, the basis for further research on how we interact with other people in the virtual environments. For example, direct perceptual acts and proprioceptive locomotion control can benefit from, and have influence on, social and psychological studies such as personal space¹, proxemic theory², and micro-territorial behaviors³. I expect that this approach will dramatically enhance the practicality of virtual world applications in many domains, including business collaboration, online education as well as other types of virtual social communities.

¹ Sommer, R. (1969). *Personal Space: The Behavioral Basis of Design*. Prentice Hall.

² Hall, E. T. (1969). *The Hidden Dimension*. Doubleday.

³ Schefflen, A. E. (1976). *Human territories: how we behave in space-time*. Prentice Hall.

REFERENCES

Literature

Bekkering et al 2002

Bekkering, H. and Neggers, F. W. (2002). Visual search is modulated by action intentions. In *Psychological Science* (13), 370-374.

Bowman et al 2001

Bowman, D. A., Kruijff, E., LaViola, J. J., Jr., and Poupyrev, I. (2001). An introduction to 3-D user interface design. In *Presence: Teleoperators and Virtual Environments*, 10(1): 96–108.

Bowman et al 2005

Bowman, D. A., Kruijff, E., LaViola Jr., J. J., and Poupyrev, I. (2005). *3D User Interfaces: Theory and Practice*. Addison Wesley.

Bradski and Kaehler 2008

Bradski, G., and Kaehler, A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media.

Bryant 1998

Bryant, D. J. (1998). Human spatial concepts reflect regularities of the physical world and human body. In *Representation and Processing of Spatial expressions*. Lawrence Erlbaum Associates Inc. pp. 215-230.

Crane 2001

Crane, T. (2001). *Elements of mind: an introduction to the philosophy of mind*. Oxford University Press, 132.

Crowell et al 1998

Crowell, J. A., Banks, M. S., Shenoy, K. V., and Andersen, R. A. (1998). Visual self-motion perception during head turns. In *Nature Neuroscience*, 1, 732-737.

Dennett 1991

Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Co.

Descartes 1641

Descartes, R. (1984 [1641]). Meditations on First Philosophy. In Cottingham, J., Stoothoff, R. and Murdoch, D. (Trans.). *The Philosophical Writings of René Descartes*, Cambridge University Press, (2): 1-62.

Descartes 1958

Descartes, R. (1958). Dioptric. In *Descartes: Philosophical Writings*. (Smith, N. K. Ed. and Trans.). New York.

Dourish 2004

Dourish, P. (2004). *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, MA, 109.

Dreyfus 1991

Dreyfus, H. L. (1991). *Being-in-the-world: A commentary on Heidegger's being and time, division I*. MIT Press, Cambridge, MA.

Dreyfus 2001

Dreyfus, H. L. (2001). *Think in Action: On the Internet*. Routledge, 55.

Faillenot et al 1997

Faillenot, I., Toni, I., Decety, J., Gregoire, M. C., and Jeannerod, M. (1997). Visual pathways for object-oriented action and object recognition: Functional anatomy with PET. In *Cerebral Cortex* (7), 77-85.

Fischer and Hoellen 2004

Fischer, M. H. and Hoellen, N. (2004). Space- and Object- Based Attention Depend on Motor Intention. In *The Journal of General Psychology*. 131(4): 365-377.

Floridi 2005

Floridi, L. (2005). The Philosophy of Presence: From Epistemic Failure to Successful Observation. In *Presence: Teleoperators and Virtual Environments*, (14) 6: 656-667.

Forsberg et al 1996

Forsberg, A., Herndon, K., and Zeleznik, R. (1996). Aperture Based Selection for Immersive Virtual Environments. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, 95-96.

Franklin and Tversky 1990

Franklin, N., and Tversky, B. (1990). Searching imagined environments. In *Journal of Experimental Psychology: General*, 119, 63-76.

Gaver et al 1995

Gaver, W. W., Smets, G., and Overbeeke, K. (1995). A Virtual Window On Media Space. In *Proceedings of the SIGCHI conference on Human factors in computing system*. 257-264.

Gibson 1972

Gibson, J. J. (1972). A theory of Direct Visual Perception, In Royce, J. and Rozenboom, W. (Eds.). *The Psychology of Knowing*. Gordon & Breach, NY, 215-240.

Gibson 1979

Gibson, J. J. (1979). *The ecological approach to visual perception*, Hillsdale, NJ.

Hall 1969

Hall, E. T. (1969). *The Hidden Dimension*. Doubleday.

Heidegger 1962

Heidegger, M. (1962). *Being and time* (Macquarrie, J. and Robinson, E. Trans.). Harper Collins, San Francisco.

Howard 1982

Howard, I. P. (1982). *Human Visual Orientation*. John Wiley, Chichester.

Jeannerod et al 1995

Jeannerod, M., Arbib, M. A., Rizzolatti, G., and Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. In *Trends in Neurosciences*. 18(7): 314-320.

Jolicoeur et al 1993

Jolicoeur, P., Ingleton, M., Bartram, L., and Booth, K. S. (1993). Top-bottom and front-behind decisions on rotated objects. In *Canadian Journal of Experimental Psychology*, 47 (4): 657-677.

Kalay 2004

Kalay, Y. E. (2004). *Architecture's New Media*. MIT Press.

Kelly 2000

Kelly, S. D. (2000). Grasping at Straws: Motor Intentionality and the Cognitive Science of Skillful Action. In *Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus (II)*. Eds. Mark Wrathall and Jeff Malpas. MIT Press, Cambridge, MA, 161-177.

Kelly 2004

Kelly, S. D. (2004). *The logic of motor intentional activity*, Draft.

Liang and Green 1994

Liang, J. and Green, M. (1994). JDCAD: A Highly Interactive 3D Modeling System. In *Computers and Graphics*, 18(4): 499-506.

Lombard and Ditton 1997

Lombard, M. and Ditton, T. (1997). A the heart of it all: The concept of presence. In *Journal of Computer-Mediated Communication*, 3(2). WWW URL: <http://jcmc.indiana.edu/vol3/issue2/lombard.html>

Merleau-Ponty 1945

Merleau-Ponty, M. (2003 [1945]). *Phenomenology of Perception*. Routledge Classics, 292.

Milner and Goodale 2006

Milner, A. D. and Goodale M. A. (2006) *The Visual Brain in Action*. Oxford University Press, NY. 126-128.

Mylov 2002

Mylov, P. (2002). On Space, Its Time, and Spatiotemporal expressions. In *Virtual Space: spatiality in Virtual Inhabited 3D Worlds*, Springer-Verlag, 47-70.

Neale 1998

Neale, D. (1998). Head Mounted Displays: Product Reviews and Related Design Considerations, Blacksburg, Dept. of Industrial Systems and Engineering, Virginia Tech, HCIL-98-02.

Noë 2002

Noë, A. (2002). Is the Visual World a Grand Illusion? In *Journal of Consciousness Studies*, 9 (5-6): 1-12.

Noë 2004

Noë, A. (2004). *Action in Perception*. MIT Press, Cambridge, MA, 59.

Noma and Miyasato 1998

Noma, H. and Miyasato, T. (1998). Design for Locomotion Interface in a Large Scale Virtual Environment-ATLAS: ATR Locomotion Interface for Active Self-Motion. In *Proceedings of the ASME Dynamic Systems and Control Division*, 111-118.

Palmer 1999

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA.

Paulignan and Jeannerod 1996

Paulignan, Y. and Jeannerod, M. (1996). Prehension movement: The visuomotor channels hypothesis revisited. In Wing, A. M., Haggrad, P. and Flanagan J. R. (eds.), *Hand and Brain*. Academic Press, NY. 265-286.

Poupyrev et al 1998

Poupyrev, I., Weghorst, S., Billinghamurst, M. and Ichikawa, T. (1998). Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. In *Computer Graphics Forum, EUROGRAPHICS Issue 17(3)*: 41-52.

Reeves and Nass 1996

Reeves, B. and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications, New York, NY.

Rheingold 1991

Rheingold, H. (1991). *Virtual Reality*. Touchstone, New York, NY.

Schefflen 1976

Schefflen, A. E. (1976). *Human territories: how we behave in space-time*. Prentice Hall.

Searle 2004

Searle, J. R. (2004). *Mind: A Brief Introduction*. Oxford University Press, 259.

Sheridan 1999

Sheridan, T. B. (1999). Descartes, Heidegger, Gibson, and God: Toward an Eclectic Ontology of Presence. In *Presence: Teleoperators and Virtual Environments*, (8) 5: 551-559.

Slater 2002

Slater, M. (2002). Course Notes on Understanding Virtual Environments: Immersion, Presence and Performance. In *ACM SIGGRAPH*, San Antonio, TX.

Sommer 1969

Sommer, R. (1969). *Personal Space: The Behavioral Basis of Design*. Prentice Hall.

Stephenson 1992

Stephenson, N. (1992). *Snow Crash*. Bantam Books.

Sutherland 1968

Sutherland, I. E. (1968). Head-Mounted Three Dimensional Display. In *Proceedings of Fall Joint Computer Conference*, 757-764.

Tan et al 2001

Tan, H. Z., Slivovsky, L. A., and Pentland, A. (2001). A Sensing Chair Using Pressure Distribution Sensors. In *IEEE/ASME Transactions on Mechatronics. IEEE/ASME*, 6(3): 261-268.

Templeman et al 1999

Templeman, J. N., Denbrook, P. S., and Sibert, L. E. (1999). Virtual Locomotion: Walking in Place through Virtual Environments. In *Presence: Teleoperators and Virtual Environments*, 8(6): 598-517.

Viola and Jones 2004

Viola, P., and Jones, M. J. (2004). Rapid Object Detection Using a Boosted Cascade of Simple Features. In *Proceedings of Computer Vision and Pattern Recognition*, 511-518.

Witmer and Singer 1999

Witmer, B. G. and Singer M. J. (1999). Measuring Presence in Virtual Environments: A Presence Questionnaire. In *Presence: Teleoperators and Virtual Environments*, 7(3): 225-240.

Zahorik and Jenison 1998

Zahorik, P. and Jenison, R. L. (1998). Presence as Being-in-the-World. In *Presence: Teleoperators and Virtual Environments*, (7) 1: 78-89.

Online Websites

CAVE Automatic Virtual Environment:

<http://www.evl.uic.edu/pape/CAVE>

CyberWalk Project:

<http://www.cyberwalk-project.org/>

George Edward Moore (Stanford Encyclopedia of Philosophy):

<http://plato.stanford.edu/entries/moore/>

Johnny Lee:

<http://www.johnnylee.net/projects/>

Matrox Electronic Systems Ltd:

<http://www.matrox.com>

NaturalPoint, Inc.:

<http://www.trackir.com/>

Oblong Industries.:

<http://www.oblong.com>

OpenCV (Open Computer Vision Library):

<http://www.opencv.org>

Tomas Reid (Stanford Encyclopedia of Philosophy):

<http://plato.stanford.edu/entries/reid/>

VirtuSphere:

<http://www.virtusphere.com/>