THE UNIVERSITY OF CALGARY

Alternate 3D Control-Display Mappings

by

Jeroen Keijser

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

December, 2007

Abstract

The desire to have intuitive, seamless 3D interaction fuels research into new 3D interaction approaches. However, research continues to rely on Brunelleschi's perspective to provide a sense of visual depth to create a 3D graphical display. The interactive control space has been mapped directly into the 3D virtual display space without much thought about the effect that perspective distortion has on the interaction.

There are a myriad of possible 3D control-display mappings. Previously, 3D interaction mapping options have focused on the manipulation of control-display ratio. In this thesis, I present a conceptual framework that provides a more general control-display description.

I conduct a user study to compare three different mappings to the commonly used mapping for 3D selection and manipulation tasks. The results indicate that all three may be considered viable alternatives. Together, this 3D control-display mapping framework and study open the door to further exploration of 3D interaction variations.

Publications

Materials, ideas, and figures from this thesis have appeared previously in the following publication:

Jeroen Keijser, Sheelagh Carpendale, Mark Hancock, and Tobias Isenberg. **Exploring 3D Interaction in Alternate Control-Display Space Mappings.** *In Proceedings of the 2nd IEEE Symposium on 3D User Interfaces (3DUI 2007, March 1011, 2007, Charlotte, North Carolina, USA)*, Los Alamitos, CA. IEEE Computer Society, pages 17-24, 2007.

Acknowledgements

My journey to complete this thesis has been long and filled with many triumphs, trials, and tribulations. I could not have completed it without the help and support from so many wonderful people.

To Sheelagh, my supervisor: Thank you for taking me under your wing. It has been my privilege to be your student. Your guidance and insight have been crucial in every league of this journey.

To Ken, my co-supervisor: Thank you for being there for me right from the beginning. Your calm and confident demeanor, support, and guidance have meant so much to me — especially in the times I doubted whether I could ever finish this journey.

To my friends in the Interactions Lab: Thank you for making me feel like the lab truly was my home away from home, and that I was part of the Ilab family. Thank you, in particular, to Uta, Marian, Rob D., Jim, Petra, Mike B., Chris C., and Stephanie for reading less than polished versions of parts of my thesis and their insightful suggestions. I feel truly blessed to have had the opportunity to meet so many wonderful and talented people. I will always remember my time in the Ilab fondly.

To Mark, Floh, and Sheelagh, my paper collaborators: Thank you for all the time you spent working with me, discussing, and informing the design of the study, helping me make sense of the results, and guiding and taking me through my first paper writing marathon.

To all the friends and colleagues I met at the Banff Centre: My time in Banff has a special place in my heart. I met so many vibrant and amazing people up on Tunnel mountain. I learnt so much, and was inspired by all of you.

To Maria Lantin: My year at the Banff Centre working right beside you was an incredible experience. You are one of the most talented, humble, kind, and fantastic people I have ever met. Thank you for allowing me to have had the honour of having you as my mentor.

To Jeff: Thank you for being there for me. Thank you for distracting me and keeping me sane, making me smile and laugh, and adding so much to my life. This journey would not have been the same without you.

To my family: Thank you for all the love and encouragement throughout my life. Thank you for supporting me throughout my thesis journey and your patience when it seemed like I might be in school forever.

To all the other people in my life who have listened, encouraged, and supported me throughout my thesis journey: I thank you from the bottom of my heart, I could not have done this without you, and I am grateful to have you in my life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

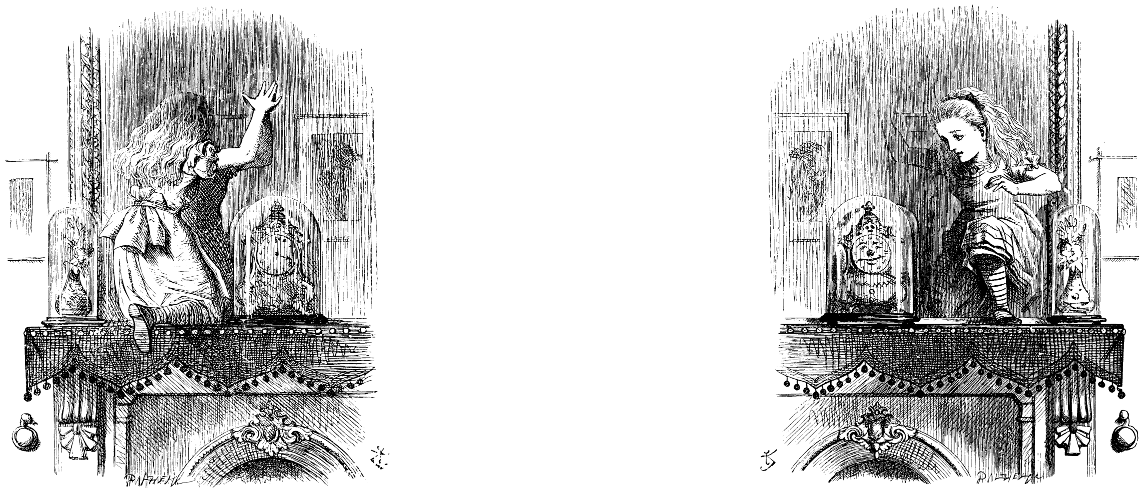
This thesis investigates the relationship or mapping between the three-dimensional (3D) Interactive Control Space and three-dimensional (3D) Display Space. This is the relationship between the physical 3D space a person (or group of people) moves and interacts in, and the virtual 3D space, displayed by the computer where a person expects to see the result of their interaction appear. While 3D graphics and 3D interaction are mature fields of research, this topic is surprisingly unexplored.

This chapter starts by motivating this investigation and setting it within the context of current 3D graphics and interaction research. Next, I present the specific thesis problems and the related thesis goals that address these problems. I finish the chapter with an overview of the rest of the thesis.

1.1 Background and Motivation

As a culture, we have had fantastical ideas about exploring alternate 3D spaces for at least a century: take, for example, "Through the Looking Glass and What Alice Found There" by Lewis Carroll [24]. In "Through the Looking Glass and What Alice Found There", Alice steps through a mirror into an alternate parallel 3D universe where she finds that ordinary inanimate objects, such as the mantle clock, come to life in extraordinary ways (Figure 1.1).

Since the advent of 3D computer graphics, people have been constructing and



(a) Alice peering into the mirror.

(b) Alice stepping through the mirror.

Figure 1.1: Alice in ‘Through the Looking Glass’ from [24].

(a) CAVE

(b) Head Mounted Display

(c) glove

(d) Wii Mote

Figure 1.2: Display configurations and Input devices

interacting with alternative 3D virtual worlds, which is particularly apparent in the gaming industry. Currently, the great variety of possible 3D interfaces — CAVEs™, 3D visualizations, 3D games — continues to proliferate. These interfaces differ not only in purpose (work or play), but also in many physical aspects, including display capabilities and physical configurations such as the CAVE™ (Figure 1.2(a)) and head mounted display (Figure 1.2(b)), and input devices such as a glove (Figure 1.2(c)) and the WiiMote (Figure 1.2(d)). Most 3D interfaces make use of the 3D projection derived from Brunelleschi's perspective [47] that has been used in computer graphics for the last forty years or more. Figure 1.3 shows this common perspective used in a game environment. However, there are some exceptions, such as incorporating perspective from the eastern tradition [25, 55, 56] or cubism inspired projection [37, 36], and recently there has been some exploration into the creation of 3D scenes with frameworks that incorporate many possible alternative projections [10, 22, 27, 51, 80]. Embellishments are being explored, such as passive [4, 20] or active stereo [1, 28] and 3D sound effects [12, 72], which can add to the power of the 3D experience, however, the basic 3D perspective still commonly stays the same.

(a) Without perspective projection. Notice how the fences appear parrallel

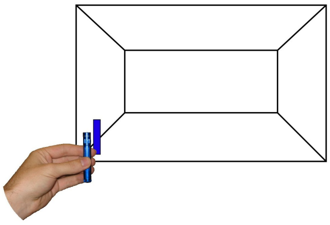
(b) With typical perspective projection. Notice how the fences and buildings on each side of the path appear to converge in the distance

Figure 1.3: Notice how the use of perspective on the right makes the world appear more realistic and more 3D (from the game called "Eternal Lands" [3])

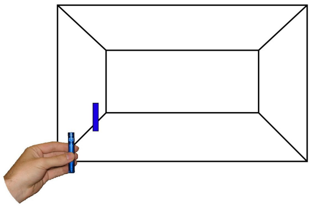
There is a significant amount of research into different aspects of 3D interfaces, such as 3D displays [28, 30, 41, 49, 71, 77], 3D input devices [7, 65, 86, 90, 91], and 3D interaction metaphors [15, 34, 58, 62, 66, 68]. In parallel, there is also a continuing discussion about and research into difficulties with 3D interfaces including problems with developing effective 3D perceptual support [17, 82], problems with developing intuitive and understandable 3D visualizations [73], and more generally discussions about lack of overall adoption of 3D interfaces [11].

One aspect that has received less research attention is the mapping of the interaction control space to the display space. To explore alternate experiences in 3D interaction, the research presented in this thesis considers the 3D to 3D mapping from a person's or people's interaction space in the real 3D world (*control space*) onto the virtual 3D *display space*. To date control to display mappings have been explored in terms of control-display gain, also referred to as control-display ratio [54, 76, 82]. These are changes in scale such as a small movement of the hand mapped to a large movement in the virtual space displayed on the screen. Beyond manipulation of scale, there has been little exploration of whether there are other viable alternatives in control-display mapping. Despite many differences amongst 3D interfaces, most explorations into 3D interaction rely primarily on perspective for the creation of the 3D display space, and commonly map the interactive control space directly into this perspective display space. This has become the default mapping, however, it does cause some perceptual discontinuities. For example (Figure 1.4), if one moves the input in control space to the lower left hand corner just in front of the screen, one's cursor in virtual 3D display spaced ends up positioned considerably to the right and higher up the screen. These perceptual discontinuities raise questions about whether this style of 3D interaction is understandable due to our ability to comprehend perspective depth cues, or whether some alternate 3D control-display mapping would be easier to work with.

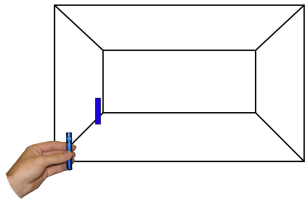
Figure 1.5 illustrates the general context and scope of my research. Within the broad area of Human Computer Interaction, this thesis research lies within 3D



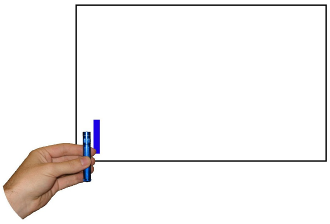
(a) Starting at the corner of the interactive space farthest from the screen, one is interacting near the front of the perspective box



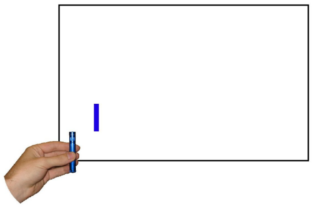
(b) Moving towards the screen to approximately halfway



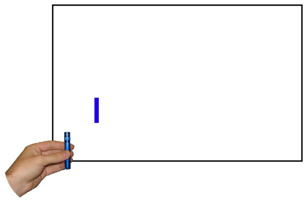
(c) Right up close to the screen, one is interacting at the back of the perspective box



(d) Starting at the corner of the interactive space farthest from the screen, the cursor appears directly in front



(e) Closer to the screen, the cursor begins to veer up and to the right



(f) Right up close to the screen, the cursor is quite far up and to the right

Figure 1.4: Mapping control space to display space: (a), (b), and (c) show a perspective display space with a perspective control-display mapping for the interaction. Note how the corresponding cursor object in display space makes sense within the box drawn in perspective but when the box cues are removed in (d), (e), and (f) this mapping is not so obvious.

Interaction. Within 3D Interaction, this research crosses the three broad areas of research: 3D Input Design, 3D Display, and 3D Interaction Techniques. Specifically, I look at 3D control-display mappings. In this thesis I develop a conceptual framework to describe many alternative 3D control-display mappings. Since the number of studies suggested by the framework is considerable, the study in this thesis considers a subset of these specifically contrasting the standard mapping to three possible alternatives.

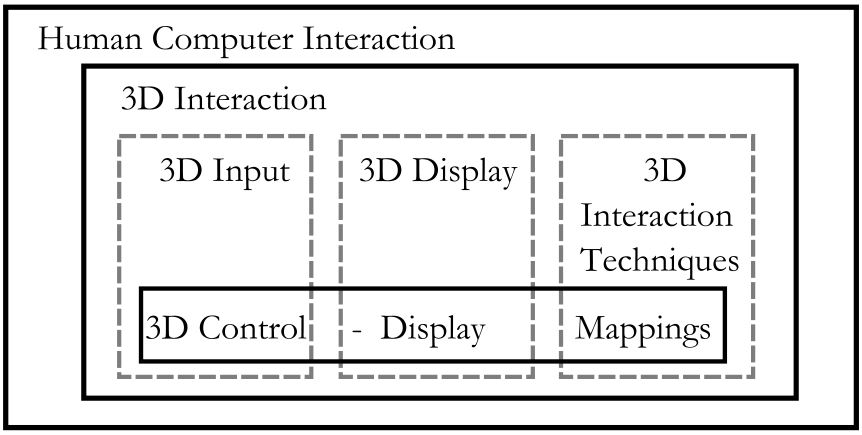


Figure 1.5: This figure illustrates where my thesis topic of 3D Control-Display Mappings lies within the larger field of Human Computer Interaction

1.2 Thesis Problems

While 3D interaction has become a well-established area of research, and improvements continue to be made, interacting in 3D is still difficult and the development of effective 3D interfaces is challenging. Many areas and ideas related to 3D control-display mappings have yet to be investigated, and to further our understanding of 3D interaction I investigate 3D control-display mappings. In this thesis I address the following problems:

1. **What are the components of 3D control-display mappings?** To date, 3D control-display mappings have only been discussed in terms of control-display gain or scale. What are the other components of this relationship?

2. **Can the components of 3D control-display mappings be formalized?** Is there a way to frame and capture the various components into an understandable cohesive whole?

3. **Are any of the alternate 3D control-display mappings viable?** Viable alternative 3D control-display mappings have not been investigated and need to be studied. We need to identify which new possible alternative mappings are worthy candidates to investigate as viable alternatives. To understand these alternatives and to indicate whether expanding our understanding of control-display mappings will have an impact on 3D interaction, empirical studies need to be performed.

1.3 Thesis Goals

I will address the aforementioned problems with the following goals:

1. **To identify what the components of 3D control-display mappings are:** I identify and describe the components of 3D control-display mappings. *(addresses Problem 1) This goal is addressed in Chapter 3 in Sections 3.1 and 3.2.*

2. **To formalize the discussion of the components of 3D control-display mappings:** I create a conceptual framework and mathematically define the components identified through Goal 1 *(addresses Problem 2) This goal is addressed in Chapter 3.*

3. **To identify and evaluate potentially viable alternate 3D control-display mappings:** I identify issues with the standard mapping and potential mapping alternatives that may address these issues. I perform an initial evaluation of the mappings by designing and running a controlled experiment. I

ask participants to perform a series of simple tasks, and to repeat such tasks with four distinct mappings (one traditional, and three alternative). I record and analyze participants' related previous experience, task performance, and mapping preferences. I further discuss and analyze the results of the study and the particular choice of the four distinct mappings *(addresses Problem 3) This goal is addressed in Chapter 3 in Section 3.3 and in Chapter 4.*

1.4 Thesis Overview

This thesis is organized into five chapters and three appendices, as follows:

Chapter 2 surveys related work in relevant perceptual cognitive science; 3D display; 3D input devices and design; 3D interaction techniques; control-display mappings; and 3D interaction evaluation. The perceptual cognitive science section discusses how we perceive our physical 3D world, and in particular focuses on depth perception. The 3D display section consists of a concise survey of all the main types of 3D displays, and a discussion of how they support 3D perception. The 3D input section discusses device factors to consider and the main types of input devices. The 3D interaction techniques section includes a 3D interaction technique task classification and which techniques are the most influenced by the 3D control-display mapping. The control-display section discusses control-display in one, two, and three dimensions. It also discussed the related concepts of isomorphism and "magic" interaction techniques. The last section on 3D interaction evaluation discusses the evaluative measures used in 3D interaction.

Chapter 3 describes the 3D Control-Display Mapping framework and delineates and formalizes these mappings *(addresses Goals 1 and 2).* This includes discussions on: the visual representation of the mappings, and the mappings themselves.

Chapter 4 describes the study I conducted *(addresses Goal 3).* Chapter 4 begins with the motivation and study design. Chapter 4 then describes the participants, apparatus, procedure, and the results. The chapter concludes with a discussion of the study and its results.

Chapter 5 summarizes my research contributions. Chapter 5 includes a discussion of future work relevant to the investigation of 3D control-display mappings and concludes with a short overall summary.

Appendix A includes study documents, such as the pre and post questionnaires and the consent form. Appendix B is the ethics approval. Appendix C is the co-author permission.

Chapter 2

Related Work

Humans interact naturally with their physical three-dimensional (3D) world. Movies such as the Matrix capture our imagination partly because we can relate to the alternate three-dimensional virtual existence they present. Unfortunately, computer-mediated virtual reality remains elusive as envisioned by futurists. However, some success has been realized in this area as testified to by military and aviation simulators. Foley [33] has successfully argued that the need for computer-mediated 3D will pay dividends in at least: Training, Therapy, and Theatre, but many more application areas exist including city planning, automotive design, and scientific visualization. All of these computer-mediated 3D application areas are dependent on 3D interaction.

Underlying all current 3D interaction work is the mapping of control space to display space. Reiterating from Chapter 1, *3D control space* is a person's interaction space in the real 3D world, and *3D display space* is the virtual 3D display space that this interaction is mapped onto. We are largely limited in the kind of interactions possible by the hardware devices available. This chapter describes the current state of the art in input (control) hardware, display hardware, and interaction techniques. This is used as the basis to frame the thesis research.

Integral to effective 3D interaction is how it incorporates our abilities to perceive 3D. To appropriately discuss the effectiveness of the input and display hardware, first, the relevant perceptual cognitive science (Section 2.1) is discussed. This chapter then frames the thesis work by reviewing the current state and limitations of input (control) hardware (Section 2.3) and display hardware (Section 2.2). This is followed by current 3D interaction techniques (Section 2.4).

After discussing 3D interaction, previous literature that pertains directly to control-display mapping (Section 2.5) is reviewed. Lastly, previous study and evaluation of 3D interaction (Section 2.6) is discussed as context to the evaluative study in this thesis (Chapter 4).

2.1 Overview of Perceptual Cognitive Science as It Pertains to 3D Interaction

To interact in 3D we rely on our senses to gain information on the spatial-temporal relationship of objects in relation to each other, and in relation to ourselves. Thus, knowledge of how we sense and interpret the 3D nature of the world is an asset in creating effective computer-mediated 3D interaction.

Of the five primary senses, sight and touch are the senses we rely on the most to interact in 3D. Visual display hardware is an ongoing area of research, however, sight is arguably the most supported of the senses. Unfortunately the fidelity of haptic display is far from the fidelity of the physical world. As such, computer-mediated touch sensations are limited to the passive haptic sensations of the input device and to the use of physical proxies — physical proxies that may or may not be of similar shape, texture and weight as the virtual objects they are proxies for. This currently relegates touch, and the use of our hands, to providing coarse input to the computer system. Sight is thus the primary sense that we rely on when interacting with computers. 2D display hardware has matured to a state that 2D displays can be found in a multitude of products and are used by the public at large. The development of 3D display hardware is far behind. Current 3D display relies heavily on 2D display hardware, and in most cases, it modifies a 2D display to display 3D (see Section 2.2). As the third dimension, or depth dimension is not inherent in 2D display hardware, 3D display is achieved through knowledge of how humans perceive it and by simulating the depth dimension. To discuss 3D display developments, the concepts of depth perception are first presented here. This is followed by a brief discussion on the proprioceptive system, our internal kinesthetic senses, as these internal senses also assist our sense of the 3D nature of the world.

2.1.1 Visual System

To perceive depth, the visual system gathers depth information based on a variety of depth cues. These depth cues can be separated into monocular cues (single eye); and binocular cues (both eyes). The monocular cues are: linear perspective, texture gradient, size gradient, occlusion, depth of focus, cast shadows, shape-from-shading, and structure-from-motion (kinetic depth, motion parallax). The binocular cues are: stereoopsis, accommodation, and convergence [85]. The following gives a brief explanation of each.

Monocular Cues

Perspective: Ware [85][1] describes *linear perspective* as follows:

> Figure 8.1 [Figure 2.1 here] shows how perspective geometry can be de-

[1] from the text "Information Visualization" on page 275

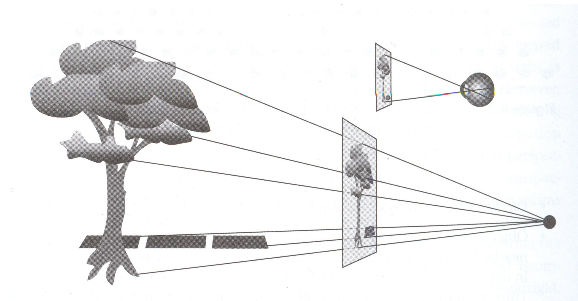


Figure 2.1: *Linear Perspective*: For each point on the canvas a ray of light is traced back from a fixed point to the objects in the scene. This has the effect that objects that are farther away appear smaller on the screen. (Figure 8.1 on page 275 in [85] used with permission)

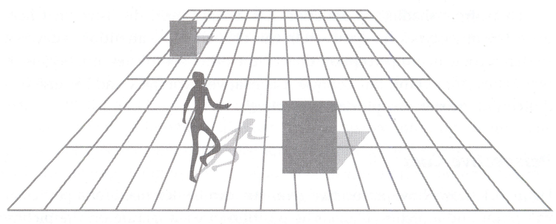


Figure 2.2: Parallel lines drawn in perspective appear to converge. A known object such as the human provides an impression of the size of the scene, and the vertical dark squares create a *size gradient*. The grid displayed can also be considered a *texture gradient*. Cast shadows indicate the squares and human figure are on the plane (Figure 8.2 on page 276 in [85] used with permission)

scribed for a particular viewpoint and picture plane. The position of each feature is determined by extending a ray from the viewpoint to that feature in the environment. If the resulting picture is subsequently scaled up or down, the correct viewpoint is specified by similar triangles as shown. If the eye is placed at the specific point with respect to the picture, the result is a correct perspective view of the scene. A number of the depth cues are a result of the geometry of perspective. These are illustrated in Figure 8.2 [Figure 2.2 here]

Texture Gradient: A *texture gradient* results from a surface with a uniform pattern or texture. The features of the pattern become smaller with distance and create a depth gradient.

Size Gradient: When a set of similar objects, such as a bag of marbles or set of cubes, are scattered over the environment, objects appear smaller the farther away they are, creating a *size gradient*. If an object of known size is added into a scene, such as a human being, it creates an impression of the overall size of the scene.

Occlusion: *Occlusion* is the result of objects overlapping each other. If object A overlaps object B, then object A appears in front of object B and is perceived as being closer.

Depth of Focus: As the eyes focus dynamically on various objects, they focus to the depth of these objects or to a particular *depth of focus*. When a particular object at a particular distance is in focus, objects farther away from and closer to the viewer are out of focus and become slightly blurred. Typically objects in a virtual scene are always in focus, and this depth cue is lost.

Cast Shadows: Lit objects cast shadows onto other objects. The size and shape of these *cast shadows*, particularly upon large surfaces, give an idea of how far apart the objects are.

Shape-From-Shading: When an object is lit, the way the different parts are lit differently gives and idea of the 3D shape of that object (i.e. *shape-from-shading* see Chapter 7 in [85]).

Motion Parallax: *Motion parallax* results from either motion of the person relative to the object or motion of the object in the environment relative to the person. As the motion occurs different sides of the object become visible and a strong notion of the 3D whole of the object is created. Further, closer objects or parts of objects appear to move faster than farther objects or more distant parts of the object which gives a relative depth cue.

Kinetic Depth Effect: The *kinetic depth effect* is an effect observed when a 3D object spins. The rotational motion of the 3D object is a depth cue that enhances the mind's understanding of the three dimensional shape of the object.

Binocular Cues

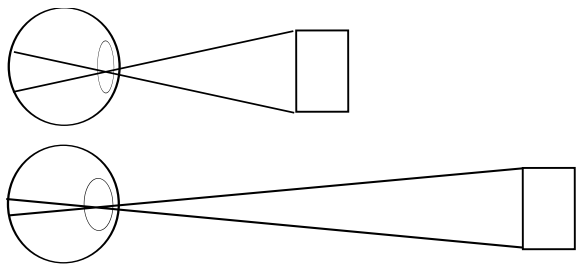


Figure 2.3: Accommodation: the lens in the top eye is more stretched to focus on the near object and the lens on the bottom eye is more relaxed and rounder to allow focus on the distant object (adapted from Figure 3.5 on page 37 in [19])

Stereopsis: *Stereopsis* or stereoscopic vision results from having two slightly different images, one in each eye, and the fusing process that occurs. To focus on an object each eye's lens adjusts or accommodates. This is called *accommodation* (see Figure 2.3). The eyes rotate slightly inward or outward or converge on the object of interest (i.e. *convergence* see Figure 2.4). *Stereopsis*, the fusing of the two images, is quite a strong depth cue and has a large impact. For some, stereoscopic support is the distinguishing feature of a 3D display over a 2D display. However *stereopsis* is only one of the many depth cues possible and most stereoscopic 3D displays have the accommodation-convergence conflict that results from displaying a stereoscopic

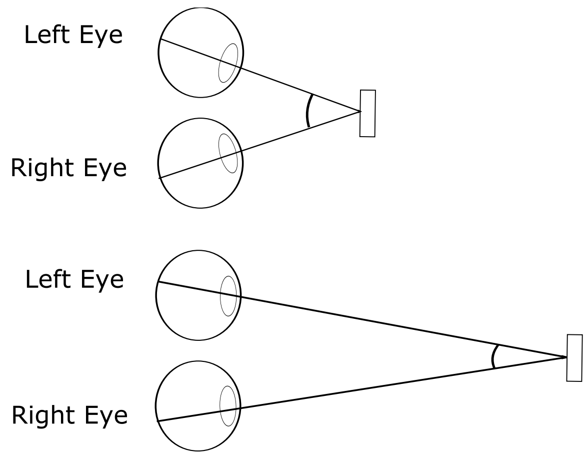


Figure 2.4: Convergence: the eyes are rotated more inward for the close object and less so for the far object. The angle created is called the vergence angle (adapted from Figure 3.5 on page 37 in [19])

3D object on a 2D display (see Figure 2.5).

Discussion on Depth Cues

Depth cues are just that — cues. Depth perception is a process that constructs depth information as a synthesis of the sometimes conflicting information that each individual cue provides. The Necker cube is a famous example of ambiguous depth cues (see Figure 2.6). As is noted by Hoskinson *et al.* [46], pragmatically, even if all depth cues currently supported are faithfully executed in high end immersive virtual reality (IVR) setups, depth perception errors remain high and varied between people. This limits the practical application of IVR to the modeling and prototyping process. Hoskinson *et al.* propose a study to determine the relative importance of the accommodation-convergence conflict to the high depth perception error rate in stereoscopic displays. Further, it is known that the amount a particular depth cue is relied upon is dependent on the particular task. Wagner *et al.* [83] ran a study

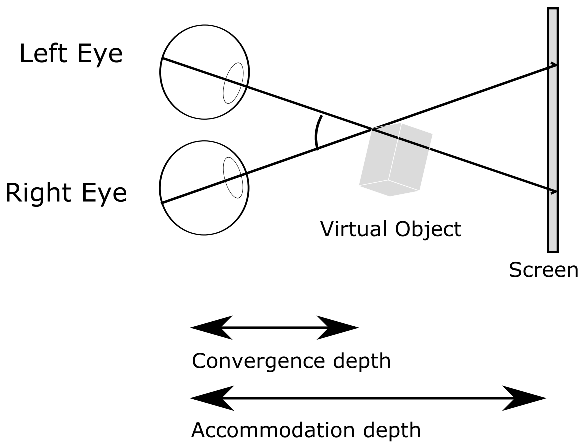


Figure 2.5: The mismatched depth information between accommodation and convergence common in many computer generated stereo systems. The eyes accommodate at the depth of the screen, and the eyes converge at the projected depth of the virtual object. This results in conflicting depth information which is referred to as the accommodation-convergence conflict (adapted from Figure 3.7 on page 39 in [19])

to compare the relative importance of depth cues in simple 3D interaction tasks of positioning an object, rotating an object, and scaling an object. Perspective projection (linear perspective) was found to generally have a positive impact on performance; however, it had a slightly negative impact on the rotation task. Cast shadows seemed to have a positive effect in all situations, but only minimally in the rotation task. The other cues tested seemed to generally be important, but did not statistically improve results.

The majority of 3D is displayed on a 2D screen (see Section 2.2). Brunelleschi, during the 15th century, had an insight into how light from the 3D world enters the eye and how you could capture that on a 2D surface, which lead to the geometrical explanation of linear perspective. Linear perspective was incorporated into Western thought, and in turn faithfully adopted by computer graphics. Though computer graphics is currently exploring beyond perspective projection [10, 27, 51], in 3D

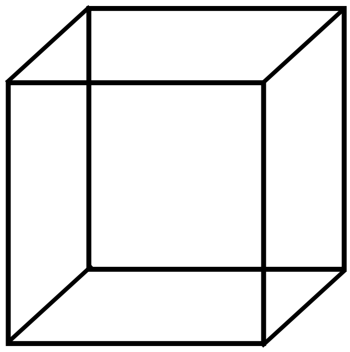


Figure 2.6: In the Necker cube shown here, it is unclear which side of the cube is in front of the other

interaction and 3D gaming linear perspective still dominates. This dominant position indicates that the implications of using linear perspective for 3D interaction should be clearly understood. This is part of the motivation for this thesis.

2.1.2 Proprioceptive System

Beyond the primary five senses, there are the senses that make up the proprioceptive system. The proprioceptive system consists of the internal senses that help us perceive movement and spatial orientation. This includes the inner ear sense of balance, but also the tensing and relaxing of our muscle and movement of our joints. These senses assist in allowing us to walk upright and are the basis for coordination. They also let us know the relative position of our various body parts in relation to each other and relative to the objects in the space immediately around the body, also known as the *peripersonal space* [69, 31].

When and how the spatial information sensed by the proprioceptive system and the visual system integrate into a cohesive whole is an area of continued research. The *two visual system hypothesis* maintains that they correspond to the two distinct models [57]. The visual system gives the sensation information for the cognitive-visual model. Likewise, the proprioceptive system gives the sensation information for the sensorimotor model. It appears that people are quite adaptable to sensory discrepancies between the two systems. Burns *et al.* [23] artificially separated the hand and eye by placing a screen directly in front of the eyes. This blocked a person's ability to see their hands. Instead the person saw a virtual representation of their hands. This allowed Burns *et al.* to separate the location of where the participant saw their hand and the actual location of the participant's hand. In the study they slowly increased the difference between virtual representation and actual location of the hand. Participants were able to successful interact, and did not notice the difference, even when it was increased to a large amount. This adaptability suggests that exploration beyond the default hand-eye relationship is possible. The control-display mapping is related to the hand-eye relationship in that the hand moves in *control space* and the eye sees the result in *display space*. Thus, this adaptability further suggests that variations in the control-display mapping, even drastic variations, may be possible.

In computer-mediated 3D, the proprioceptive senses are those that we rely on when interacting in the *control space* and visual system and depth cues are used to sense the *display space*. The depth cues discussed above are used to discuss the effectiveness and limitations of display hardware in the next section.

2.2 3D Display

To date, 3D interaction (3DI) research, particularly in virtual reality research, has a large focus on stereoscopic display. However the majority of current stereoscopic displays introduce artificial depth cue conflicts (i.e. the accommodation-convergence conflict as described in Figure 2.5) which may seriously diminish the effectiveness of using stereoscopic display broadly [46]. Further, current 3D display technology, as purely information displays, only allows the addition of 3n or 3 times as much information as on a 2D display [87].

The ability to interact with what is displayed is as important as the display itself. Unfortunately, as mentioned in the previous section, the hardware to support the haptic channel is severely limited. This has created a distinctive separation between input and display. This distinctive divide is the basis for a *control space* and separate *display space* and this divide must be acknowledged and its implications understood. Further, to discuss the control-display mapping, it is important to understand the implications of the physical hardware on both the display and control side. 3D display hardware is discussed in this section and 3D input (control) hardware is discussed in Section 2.3.

For displaying 3D currently, there are three main approaches: displaying on a 2D flat screen, adding a few pieces of hardware to use a 2D screen as a stereoscopic screen with binocular depth cues, and creating new displays that are closer to being a 3D display that matches 3D in the physical world (volumetric displays and autostereoscopic displays). Each approach is discussed in turn.

2.2.1 Flat Screens with Monocular Depth Cues

CRT monitors, LCD monitors, large tiled wall displays (see Figure 2.7), and projectors, are all effective 2D displays. The majority of monocular depth cues can be displayed directly on a 2D display, and the success of 3D computer animation in the gaming and movie industry illustrates that 3D can be displayed fairly effectively on a 2D display. Of the monocular depth cues discussed in Section 2.1.1, only dynamic *depth of focus* and *motion parallax* cannot be directly supported on a 2D display. Dynamic *depth of focus* is computationally expensive [59] and is rarely supported (notable exceptions [59, 70]). Current implementations require eye tracking to determine which pixel a person is looking at, and then use the determined pixel to read depth information from the associated z-buffer to ascertain the depth a person is looking at [70]. Then, various computer graphic blurring techniques can be applied to areas outside the determined *depth of focus* to simulate the effect of this depth cue. *Motion parallax* requires head tracking data to move the visual space relative to a person's head. The addition of *motion parallax* through the use of head tracking comes with a strong caveat. It is effective for a single person, as the visual space moves relative to a single person's head. When displayed to an audience, the whole visual space still moves according to the single tracked person's point of view, and the lack of control can induce cybersickness in the audience. Thus, it is possible to support *motion parallax* on a flat screen; however, it is best suited for single user applications only.

Figure 2.7: A tiled wall display is an example of a flat screen

2.2.2 The Addition of Binocular Depth Cues

All of the discussed binocular depth cues can not be displayed directly in a 2D display. To date, 3D display that equals what we see in the physical world is not technically feasible (though 3D display development is ongoing, and each new display claims to have achieved this with names such as Actuality™ [2], Real-depth 3D™ [5], or SeeReal™ [6]) and the third dimension on a 2D display is faked in one manner or another. The first approach is to extend a 2D display to create a stereoscopic display by providing a different image to each eye with the addition of shutter glasses, polarized glasses, or even the use of separate displays for each eye.

One of the earliest stereoscopic displays consisted of shutter glasses with a regular CRT monitor and half-silvered mirror that allowed the control space and display space to overlap [71]. Shutter glasses work by alternately blocking light to the one or the other eye. The CRT monitor must correspondingly alternate between two images rapidly in synchronization with the shutter glasses to create the stereoscopic effect. This is known as active stereoscopic display. Passive stereoscopic display consists of the use of polarized light to display two different images at the same time. The left and right image are displayed with light oppositely polarized. The polarized filter in each eye of the glasses allows only light polarized appropriately to enter the respective eye. Circular polarization is generally preferred over linear polarization as it allows a person to tilt their head. Passive stereoscopic displays are generally preferred as they are less likely to induce headaches or nausea [9] caused by the flicker inherent in active stereoscopic displays.

Head Mounted Displays

Figure 2.8: An example of a head mounted display

Head mounted displays (HMD) (see Figure 2.8) are an example of the separate display for each eye approach. Early head mounted displays were worn on the head and completely blocked out light from the physical world. This complete separation from the physical world was thought to allow a person to immerse more completely within the virtual world. HMDs had been, and often continue to be, quite heavy and create a fair amount of fatigue. HMDs also suffer from a small field of view, increasing the need for head movement, and thus further fatigue. Modern HMDs have become quite lightweight, and with see-through HMDs it is possible to augment the physical world, and create mobile applications.

CAVE™

The CAVE™ (see Figure 2.9) was introduced in the early 1990s [28] as an alternative to HMDs. To increase the immersion potential, large screens surround a person who wears only stereo glasses (either shuttered or polarized) and small tracking devices. CAVE™ installations are large and fixed: a large space, often a whole room, must be dedicated to CAVE™ applications. Further, CAVE™ systems usually have relatively low resolution for the size of the screens. Although the size of the space would seem to indicate that multiple people can collaboratively interact with the system, the CAVE™ system uses head tracking to allow for properly oriented perspective, as well as *motion parallax*; and, as mentioned above, the addition of *motion parallax* through head tracking effectively means the CAVE™ system is a single user setup.

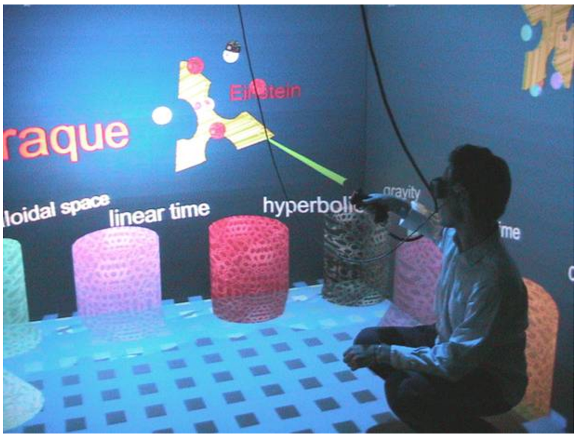


Figure 2.9: An example of a CAVE™ system

2.2.3 Volumetric Displays

An alternative to the above approaches is the volumetric display. Volumetric displays spin a 2D screen within a volume, and thus project light into a 3D volume. Therefore, they do not suffer from the accommodation-convergence conflict [41]. Also volumetric displays can be viewed from any angle and thus allow multiple people to view, point out features, and discuss a virtual 3D object, in a similar manner as gathering around an actual physical 3D object. Developing appropriate interaction techniques for volumetric displays is an ongoing area of research [42, 43]. The main detracting feature of volumetric displays is the 3D volume is encased and thus you cannot interact directly with virtual objects in the volume. Current implementations also create a distracting amount of sound (as observed in practice) and do not have a consistent image quality throughout the volume nor sufficient resolution [14], but as the hardware improves these issues are likely to be minimized.

2.2.4 Autostereoscopic Displays

Another 3D display alternative is autostereoscopic displays [30]. Autostereoscopic displays display stereoscopic images without the need for special glasses. Of the three types of autostereoscopic displays Dodgson [30] describes – parallax barrier, lenticular, and integral imaging – integral imaging with its parallax in all directions has garnered a great deal of attention [60]. Kobayashi *et al.* [49] discuss how integral imaging allows one to view a scene from as many angles as are supported in both horizontal and vertical directions along a vertical display. This is beneficial in that it allows multiple people to view the stereoscopic image, and allows for some motion parallax without head tracking. Further, it allows the display to be used as a horizontal or tabletop display. Lenticular and integral imaging displays suffer from color moire, a rainbow like effect, but recently Kobayashi *et al.* [49] were able to overcome this.

2.2.5 Discussion on 3D Displays

None of these types of 3D display offer 3D display that matches our physical 3D world, although some are coming closer, such as volumetric displays and integral imaging. When comparing CAVE™s, Wall displays, and desktop displays; none of the displays have any particular advantages in terms of navigation tasks [77] (see Section 2.4.3). The majority of 3D applications continue to be displayed on regular 2D screens, and in most cases, perspective projection is used throughout, and as such this is used in this thesis's study.

The next section discusses the other main hardware component in 3D interaction, namely, input (control) hardware.

2.3 3D Input

To interact in 3D is more than simply viewing an exquisitely rendered 3D scene. The 3D display possibilities were discussed above. However, with each possibility the depicted 3D objects and scenes have a visual appearance but no physical existence. People cannot simply grasp, squish, mold, touch, or manipulate objects as they would with their hands in the physical 3D world. Instead, to interact with these virtual objects, one uses a 3D input device in *control space*. The input device data is then mapped onto the *display space* to interact with these virtual objects.

This distinct separation between the virtual objects in *display space* and the input hardware used to interact with them in *control space* is why the 3D control-display mapping underlies all current 3D interaction.

It is possible to use conventional 2D input devices such as the mouse and keyboard to interact with these virtual 3D objects and scenes, but a plethora of 3D input devices have been developed (see 3DUI bibliography [65]) to try and capture the power and strength of our 3D capacity in the real world. To discuss the main 3D input device types (Section 2.3.2) effectively, device factors, that is, the attributes of an input device that impact its usage, are discussed first. These device factors allow one to decide which input device is appropriate for a particular application. In this thesis, these device factors were used to select the input device used for the study in Chapter 4. Further, some of these device factors are significant in the discussion of control-display mapping in Chapter 3.

2.3.1 Input Device Affordances and Factors

In the choice of input device or devices, each type of input hardware has varying affordances or factors to take into account. Zhai [91] discusses a number of device factors: *form factor, device persistence, degree integration, and isotonic versus isometric.* The following discussion includes Zhai's factors and adds *input fidelity, absolute versus relative input, tethered versus wireless, and type of tracking technology.* These are discussed and defined in order to discuss the effectiveness of input devices in Section 2.3.2, and later effectively discuss the impact of input devices in the control-display mapping.

Form Factor: As with all tools, the physical form of the tool can either impede, enhance, or suggest interaction. For example, a ball or sphere suggests rotation, and indeed rotation is quite simple to do. However, as the sphere has no indication of what orientation you started in and so the sphere would be a difficult input device to use if precise control of the amount of rotation from the original orientation is desired. This effect can be mitigated by adding virtual visual cues. It may be further mitigated if the virtual object itself has a reference point (*e. g.* the point on an arrow or cone). When choosing an input device for a particular task, one should keep in mind whether the *form factor* impedes or supports the desired task.

Device Persistence: The mouse is a device that persists at the last interaction point. One can let go of the mouse and at any future point in time can immediately acquire that last position and continue working. With free floating interaction devices this is not the case. If you wish to use your hands for some other task, you must first set the device down and in the process the last point of interaction is lost and returning to it can be difficult. This is an important factor to consider in tasks where precise interaction is important such as surgery, or any task that requires a quick and rapid change between manipulation and discussion with a colleague.

Degree Integration: *Degree integration* refers to the integration of degrees of freedom. For example, people have difficulty separating rotation into rotations around the three individual axes [61]. Devices that integrate all the supported degrees of freedom are therefore generally preferred.

Isotonic versus Isometric: An *isotonic* device is a free moving device where the position of the device is measured and used as input. An *isometric* device, alternatively, measures the amount of force applied. *Isometric* devices, such as the joystick, return to an original or nulling point if a force is no longer applied. *Isotonic* free moving devices are typically easier to learn [91], however *isometric* devices allow minimal movement to interact in a full space. However, the need for sustained pressure can increase fatigue.

Input Fidelity: *Input fidelity*, in this context, consists of device response rate, as well as tracking precision. For effective 3D interaction, the minimum input device response rate is approximately 20 hertz or 20 updates per second [86]. Tracking precision is the measure of how accurately a system can track objects. High end tracking systems are typically in the centimetre or millimetre precision range. There often exists a trade off between device response rate and tracking precision. This is typical in visual based tracking systems, where a highly precise tracking system is possible, often, at the expense of introducing an unsatisfactory degradation in response rate.

Absolute versus Relative: An *absolute* input device when located at the same physical location (or orientation) results in the exact same control space location (or orientation). This is then mapped into the display space based on the control-display mapping. *Relative* input devices return information that is relative to the last position or orientation. The typical 2D stylus is a 2D example of an *absolute* input device and the mouse is a 2D example of a *relative* input device.

Tethered versus Wireless:

Traditionally, it was technically simpler and faster to send data over a wire. Most input devices had a cable attached to the computer, and in effect tethering them to the computer. The dangling cable(s) can disrupt the sense of immersion and presence. It is common with HMD systems for an assistant to walk behind the user and carry the cables to prevent the user from tripping, which is a safety issue as well as a disruption to presence. Wireless options are preferable, but they are an expensive upgrade in most commercial systems.

Type of Tracking Technology:

The three main types of tracking technology are based on detecting magnetic, audio, or light waves.

Magnetic tracking technology was one of the earliest tracking technologies adopted by the VR community. The Polhemus® tracking system is an example, but it is notoriously susceptible to interference by metallic objects in the room including something as innocuous as jewelry worn by a person. Current high end magnetic systems such as Ascension's Flock of Birds® are more robust, but also suffer from interference from metallic objects.

Acoustic tracking, such as the Intersense IS-900®, typically place ultra-sonic emitters on the ceiling while a person wears a detector. This dedicates the ceiling to this emitting technology (*i. e.* one cannot place a display surface on the ceiling for a full CAVE). This is because unlike magnetic tracking, acoustic tracking can not work effectively through walls and screens. However, acoustic detection sensors have 180 degree detection which is a larger angle of detection than vision.

Vision systems work either with the camera/detector in the input device, like the Nintendo's Wiimote, or with cameras installed in the environment. Vision systems work by detecting objects or body position from raw images, sometimes aided by visual markers, such as Vicon®, or infrared emitters placed on the object or person's clothing especially near joints, digits, limbs, top of the head *etc.* Vision based tracking has gained popularity recently as it is possible to construct a tracking system with commodity digital video camera and computer hardware.

2.3.2 Types of Input Devices

Zhai [91] distinguishes between three main types of input devices: “flying” mice, “desktop” devices, and multiple-degrees-of-freedom (multi-DOF) armatures.

“Flying” mice, as the name suggests, are free floating interaction devices, such as the wand, that are in essence an attempt to replicate the mouse as a 3D input device. These devices are isotonic absolute input devices that are typically quite easy to learn and easy to use efficiently [90]. Although, because they are absolute input devices, they have a fixed input range. To increase the range of interaction, clutching techniques are introduced; such as pressing a button to disengage the motion of the input device from motion in the display and then moving the input device back across the space before reengaging. This has the effect of making the input device a relative input device. As “flying” mice lack device persistence, they are less suitable for tasks that require a high degree of precision.

“Desktop” devices are mounted or set on a stationary surface. These devices are typically self-centering, either because they are isometric or have some elastic property due to being spring-loaded. One of the benefits of being set on a desktop surface is device persistence as it is easy to continue again after a break exactly where one left off. However desktop devices have a longer learning phase than “flying” mice [90].

Multi-DOF armatures are a less common set of 3D input devices that use a mechanical arm to suspend the interaction device in the air. Other than the challenges that arising because some positions are not possible due to the mechanics of the armature, interacting with the armature is isotonic and thus easy to learn. Multi-

DOF armatures do not have to deal with tracking interference, and have high input fidelity. Depending on armature's type it may be a persistent device as well. However due to the physical armature, it can be awkward and tiring to use in certain positions [91].

2.4 3D Interaction Techniques

This section discusses 3D interaction techniques relevant to this research. I follow Bowman *et al.*'s [19] task classification to broadly categorize techniques as they are applied to tasks in four areas: system control; symbolic input; navigation; and selection and manipulation. Each is discussed in turn below, with a more in-depth discussion of the most directly relevant area, namely: selection and manipulation.

2.4.1 System Control

Bowman's first broad task area, which applies to any computer system, is system control. No matter how well designed a particular application is, at some point a person will still need to deal with the underlying system and do such tasks as save a sketch, load a model, replay a scene, or switch to another application. These type of tasks are called system control tasks. Too often the default is to overload a 3D input device and create some sort of "window, icon, menu, pointer" (WIMP) style menu as a catch all for system control. In contrast, BLUI [21] is an example that clearly separates system control from actual 3D interaction. In BLUI, system control is done gesturally, whereas the 3D sculpting is done with the 3D input device. Generally, system control or interacting with the underlying computer system is separate from interacting in the virtual world. However, it may be indirectly affected by the control-display mapping if it is implemented as interaction within the virtual world such as floating 3D menus. Readers interested in system control should see Chapter 8 in 3DUI [19].

2.4.2 Symbolic Input

Symbolic input is the second task area from Bowman's classification. Symbolic input is the task of communicating symbolic information (text, numbers, scribbles, and other symbols) to the computer system. This task can be accomplished in a multitude of ways. For example, it can be done through gesture, through speech, or with a keyboard. None of these methods are effected by the control-display mapping. Most methods are not affected by the control display mapping, and thus symbolic input is only discussed briefly here. Symbolic input is important in our interaction with a computer system as it is important to be able to annotate what we have done and accomplished for us or for others to be able to go back to the work later. The keyboard tends to remain as the fall back input device, but obviously is ill suited for many 3D applications where a person is walking around freely. There are many other methods of symbolic input, some with considerable promise for 3D interaction applications such as the predictive pointing text entry system Dasher [84] and mobile phone text entry. However as with system control, symbolic control is not generally affected by the control display mapping.

2.4.3 Navigation

Navigation consists of two parts: wayfinding — determining where you are and how to get where you want to go; and travel — the actual method of getting there. Wayfinding is a cognitive task. As such it is not directly effected by the control-display mapping. However, if a 3D visual aide is created to assist in this task, interacting with this visual aide will be affected by the control-display mapping. Travel involves movement through the virtual world and thus is affected by the control-display mapping. Typically an input device is overloaded with travel, selection, and manipulation. In fact, travel is often accomplished through a variation of one of the main selection techniques. For example, one of the simplest and most common traveling techniques is to point in a direction and then invoke an additional command by pressing a button, making a gesture, or using a voice command. The control-display mapping directly impacts this method as the person is in control space and points in a direction in control space that is then mapped into display space. Other navigational techniques explicitly use the control-display mapping such as World In Miniature (WIM) [76]. WIM is a technique to allow a person to interact at two different scales, and thus two different control-display mappings, one locally and the other in the global overview as represented by the WIM. Although, navigation is affected by the control-display mapping or directly manipulates it, selection and manipulation are the most directly affected and are thus discussed in more detail next.

2.4.4 Selection and Manipulation

The last task area, selection and manipulation, is the most directly affected by the control-display mapping because most techniques involve motion in *control space* mapped into *display space*. Selection and manipulation are discussed together as the act of selecting an object to interact with is the precursor to manipulating that object. Although manipulation is becoming a broader notion, as more sophisticated interaction techniques are developed, such as techniques to deform objects [35]; here, manipulation consists simply of rotating and positioning the selected object. Poupyrev and Ichikawa [67] introduced a taxonomy of selection and manipulation techniques (see Figure 2.10) divided into exocentric and egocentric techniques, where egocentric is further separated into virtual hand and virtual pointer. The exocentric techniques relate objects and places relative to each other within a broad mental map. The egocentric techniques relate objects and places relative to a person's body. The authors make an important metaphorical distinction between:

a) like *touching* an object (which they call virtual hand).

b) like *pointing* to an object (which they call virtual pointer).

One of the goals in 3DI is for a person to feel completely present in the virtual world. Thus the distinction above is based on the premise: *one is in the virtual world* and can act in it *like* one does in the real physical world. One can touch objects *within reach* and point to objects *at a distance.*

There are several issues with this taxonomy. The first is Poupyrev and Ichikawa's [67] use of the terms virtual hand and virtual pointer to label the two distinct interaction styles. Interaction techniques that use the *within reach* style do not all have a

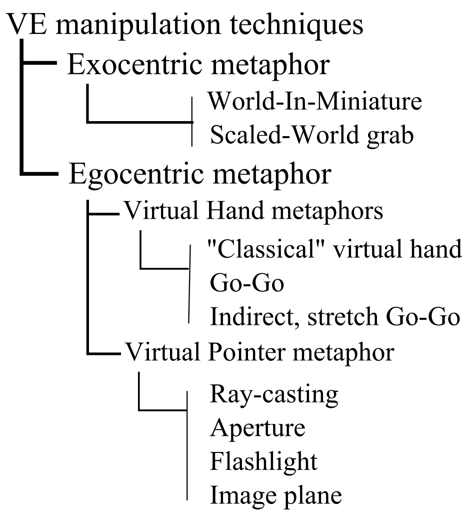


Figure 2.10: Poupyrev and Ichikawa's Interaction Taxonomy (adapted from Figure 3 on page 25 in [67]). VE is an abbreviation for virtual environment

hand as the cursor representation nor does one always interact as if one could grasp the objects with a virtual hand. The term virtual hand is misleading as it does not account for techniques that do not use a virtual hand cursor representation nor interaction style. These include precise single point cursors, point volume cursors, and direct representations of the input device itself. The term virtual pointer is ambiguous. The taxonomy authors are referring to interaction techniques that use the *action* of pointing and not a cursor point — the precise single point cursor representation. Third, when Mine *et al.* [58] introduced the scaled-world-grab technique, they argued for the use of proprioception and doing everything relative to one's own body, *i. e.*, in a egocentric fashion. Fourth the go-go hand (see 2.5.3) and extended go-go hand break with the metaphor of touching by extending the reach of a person beyond what a person could normally touch in the physical world, and so should probably be in a separate class. Lastly, the *one is in the virtual world* premise leads one to ignore the fact that the person is in a physical interactive space distinct from the virtual space.

The taxonomy is thus perhaps not as robust as one would prefer, but the distinction outlined is referred to regularly. Bowman *et al.* [19] refer to this metaphorical distinction and call a) direct manipulation and b) interaction by pointing. Direct manipulation is also an overloaded term, and as such a) will be referred to as the *within reach* metaphor and b) as the *at a distance* metaphor.

2.5 Control-Display Mappings

The mapping between the 3D control space and 3D display space has not been a common research topic. However the distinct separation between control space and display space has been known [71] since the early explorations into IVR. In 1983, Schmandt [71] discusses his attempt to have the control space and display space occupy the same physical space, which he calls spatial input/display correspondence, by the use of a half-silvered mirror. He discovered that improper occlusion of the hand by the 3D image, when the hand should be in front of the 3D object, detrimentally affects the realism of the virtual 3D scene. Improper occlusion of a person's hand by virtual objects, and vice versa, continues to be an issue with projected 3D displays; and recently Lemmerman and LaViola [50] suggest that manipulated objects should be slightly offset to avoid this occlusion.

Today there is little discussion of the 3D control-display mapping, but isomorphism (Section 2.5.2) and magic techniques (Section 2.5.3) are sufficiently related to warrant discussion here. Before these are discussed, related 2D control-display mapping literature is presented.

2.5.1 2D Control-Display Mappings

The study of the relationship between control and display originates early in the previous century with the introduction of technology that separated the action from its effect. Knobs, dials, and sliders were introduced to control things that could no longer be directly seen, and thus displays were introduced to convey what was happening. The control-display mapping was discussed with two components: gain (*i. e.* scale), and order [48]. All these controls were one dimensional, and so gain was simply the ratio between how much motion in control was transferred to the effect or motion in the display. Order, on the other hand, was whether the control affected the position, velocity or acceleration in the display.

The study of control-display gain [54] (*i. e.* scale), also referred to as the control-display ratio [13], was incorporated into 2D HCI quantitative studies. In the HCI community there is a long tradition of conducting Fitt's Law based studies and introducing techniques that improve target acquisition, that is, selection. There are a number of approaches to do this: use expanding targets as a person approaches them, or conversely increasing the cursor size, like the Bubble Cursor [40] (recently implemented in 3D [81]). An alternative approach, discussed in Semantic Pointing [13], is to dynamically change the control-display ratio to allow more control around semantically important components in an interface. This is similar to the PRISM [34] technique discussed below, which applies the same principle to manipulation instead of selection.

Other studies hinted at the possibility that there is more to control-display mapping than simply scale and order. One such study looks at the relative orientation between the 2D control space and the 2D display space [88]. Extending this study of control-display mapping to 2D and then 3D suggests that there are potentially more components to this relationship than just scale and order (as the 2D orientation study [88] implies). The next chapter delineates what these other components are, and formalizes them into a framework.

2.5.2 Isomorphism

There is a train of thought in 3DI, particularly in simulation and training, that desires to keep 3D interaction as a direct physical correlation or one to one mapping of motion. This is called *isomorphism*. Closer investigation shows that the default direct mapping is likely not a true one-to-one mapping as it ignores the issue of perspective distortion which leads directly to this thesis work.

The interaction techniques that are considered isomorphic are those that follow the basic *within reach* metaphor such as the simple or "classic" virtual hand, single point cursor, and volume cursor [92]. The extensions (go-go [66] *etc.*) break from reality, and are discussed in the next section on magic techniques. The *at a distance* interaction techniques are also not considered isomorphic as in the physical world we do not have Harry Potter's talents with a wand and people can not point at objects and levitate them. These latter techniques do not change the 3D control-display mapping and are thus not discussed here. Interested readers should see Section 5.4.2 starting on page 150 in 3DUI [19].

2.5.3 Magic

There is an opposing viewpoint in 3DI that argues virtual reality is virtual and does not have to be directly correlated to the physical world. It is argued that not sticking to what is physically possible and going beyond into what some call the "magic" interaction techniques [18] opens up a whole new realm of interaction techniques and provides the power and potential to create more effective 3D interaction. The "magic" interaction techniques discussed here are those that change the control-display mapping. All of them manipulate the scale component of the control-display mapping.

Go-Go: The Go-go interaction technique [66] is based on the popular cartoon "Inspector Gadget" who had many different gadgets including gadget arms. To engage them he would say, "Go go gadget arms". These gadget arms extended his reach beyond what he would normally be able to reach. Poupyrev *et al.*'s [66] implementation of this technique dynamically changes the scale of a person's reach in a non-linear fashion. Within two thirds length of a person's reach, the system assumed the person wanted to manipulate objects close at hand and did not change the scale. In the last third, when a person is extending their arm, it extended the reach nonlinearly depending on how much the person extended. Similar to Inspector Gadget's arms, this technique increases the reach dramatically, but does not always allow a person to grasp every object in view as it is limited to the extent a person can reach. The non-linear mapping is still a positional mapping. It simply gives the person longer arms.

Voodoo Dolls: The voodoo dolls interaction technique [62] is a two-handed interaction technique that allows a person to readily interact with objects at multiple scales. Voodoo dolls uses the non-dominant hand as a reference frame for interaction done by the dominant hand as is suggested by Guiard and Hinckley *et al.* [44, 45]. Selection is done by aiming a crosshair at the desired virtual object and pinching the virtual object between the thumb and forefinger[2]. Once the virtual object is selected a duplicate doll is created and scaled to a size that is easy to manipulate. Thus, this technique allows dynamic change of scale in the control-display mapping to facilitate manipulation.

Scaled-World-Grab: The scaled-world-grab technique [58] is similar to the go-go technique in that it extends the reach of a person. The method differs in that when a person extends their arm, the entire virtual world interactively scales around the person so that the first virtual object along the vector from the head to the hand comes within reach of the person. This is an explicit change in the control-display mapping to allow a person to manipulate any object in view.

PRISM: Precise and Rapid Interaction through Scaled Manipulation (PRISM) [34] discusses the idea that a person would like to be able to do rough large movements and small precise manipulations of an object naturally and intuitively. This technique is the opposite of the above techniques as this technique decreases the control-display ratio, *i. e.* scale, to increase precision. To determine whether a person wishes to do rough large scale manipulation or precise manipulation, the system takes advantage of the fact that people often quickly make large motions and slow down to do precise and minute manipulations. Thus, during quick motions the system is a standard one-to-one scale, but with slow motions the scale is reduced so that a person's motion

[2]This is like the "I'm crushing your head!" game that originated in a Kids in the Hall skit [8].

maps to a smaller motion in the virtual world.

Non-Isomorphic 3D Rotation: Poupyrev *et al.* [68] is notable because it explicitly refers to isomorphism and the control-display mapping. They introduce the first techniques that use rotational scaling. They argue that rotational scaling could potentially be useful to overcome rotation limitations of the human arm and/or an input device that does not allow full 360 degree rotation. Rotational scaling differs from positional scaling in that *absolute* versus *relative* input results in potentially different rotations. With a rigorous mathematical explanation, the authors conclude that *relative* input must be used for rotational scaling to be effective. They also discuss the concept of rotational *order*, another component of the 3D control-display mapping, but do not discuss or experiment with it.

All of these "magic" interaction techniques only change the scale component of the 3D control-display mapping, and only the last method refers to another component of the 3D control-display mapping. In the next chapter, this thesis introduces the other components of the 3D control-display mapping.

2.6 Studying 3D Interaction

There have been numerous 3D interaction studies in the areas of 3D display, 3D input, and 3D interaction techniques: studying 3D perceptual support of 3D displays, studying the quality and usefulness of 3D input devices, and studying the effectiveness of 3D interaction techniques. The most relevant results from those studies have been discussed through this chapter. However, the methodology and evaluative measures used have not been discussed. The relevant methodological issues are discussed in Chapter 4. The evaluative measures are discussed here.

2.6.1 Discussion of Evaluative Measures

Boritz [17] suggests interaction with a computer system should be like successful and effective conversation. He adapted Grice's [38] maxims of conversation to the evaluation of 3D interaction: these adapted maxims of conversation are an interesting and valid starting point. Boritz's four adapted maxims are defined as follows:

1) **speed** – the input action can be performed quickly
2) **accuracy** – the input action matches what is required
3) **conciseness** – the input action contains only the information needed
4) **felicity** – the input action does not put undue physical, mental, or emotional strain upon the user

In human motor capacity there is a trade-off between speed and accuracy: a person can achieve high accuracy slowly, or complete a task quickly with less accuracy. This trade off is captured in Fitt's law. Fitt's law is generally well accepted within the HCI community. Fitt's original experimental results discussed a one dimensional tapping task; however, due to its popularity it has been extended to 2D [52] and 3D [39]. Unsurprisingly, speed and accuracy are common metrics in many HCI evaluation studies. Felicity, or ease of use, is also often discussed. However, its implications on speed and accuracy are not always fully considered. To illustrate the effect felicity can have on speed and accuracy, Boritz discusses Chung's head-operated beam targeting [26]. The task in Chung's study was to find a beam direction that cut through a tumor while going through as little surrounding brain matter. The result of the study indicated that participants avoided positions that created a large amount of strain on the neck, particularly looking straight up or down, even when this would be the most effective and accurate beam direction.

Boritz continues with the addition of the maxim of *conciseness*. He cites “Put That There” [15] as an example of a project that increased speed, accuracy, and felicity at the cost of conciseness. In “Put That There”, a six DOF input device and speech commands are used to do what is in essence a 2D task of placing objects on a map. Concise communication becomes important when the same idea needs to repeatedly communicated over and over again. It is common for people to develop shorthand. For example, in written and verbal communication, experts in a certain field develop and use jargon to communicate complex ideas concisely. The use of jargon is concise, accurate, quick, and easy to use for the experts that have learnt it.

Ease of learning is a measure that is not directly included in the four maxims, and one that Zhai [91] discusses. Zhai’s evaluative criteria are similar to Boritz’s adapted maxims. He includes speed, accuracy, fatigue (part of felicity), and adds ease of learning. Ease of learning is a particularly important criterion to keep in mind with the introduction of jargon. In 3DI, examples of jargon include concise instructions such as shortcut button combinations and sets of gestures. The overhead of learning any jargon can be significant. Thus, ease of learning is the fifth maxim that is balanced against the other four maxims in determining the appropriateness of any interaction technique for a particular application.

All of these maxims are considered in the study in chapter 4. Although, as is traditional in HCI, there is more focus on speed and accuracy.

Chapter 3

A Conceptual Framework for 3D Control-Display Space Mappings

To be able to discuss 3D control-display mappings in a consistent way, this chapter formalizes the discussion by introducing a general framework and mathematical description that can be used to model and understand such mappings between control and display space. First, 3D control space, 3D display space, and the mapping relationship are defined. The framework is then presented as a whole, and each component is defined and discussed in turn. The chapter finally addresses the flip and skew mappings that are used for the study discussed in the next chapter.

3.1 Defining 3D Control-Display Space Mappings

To discuss 3D control-display mappings, first *3D control space* and *3D display space* must be defined. Reiterating from Chapter 1, *3D control space* is a person's or people's interaction space in the real 3D world, and *3D display space* is the virtual 3D display space that this interaction is mapped onto.

A 3D control space can be mapped to a 3D display space in many different ways. In general, a control-display space mapping can be described as a function, F, that maps the change in input degrees of freedom (DOF) in control space to a change in display degrees of freedom in display space:

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^m \tag{3.1}$$

where n is the DOF of the control space and m is the DOF of the display space. This varies highly depending on the particular 3D input device capabilities (from 2DOF such as a mouse, to 6DOF such as a wand).

To begin, lets discuss six degrees of freedom (three positional, and three rotational) in both control and display space, and consider mappings of the form:

$$F : \mathbb{R}^6 \rightarrow \mathbb{R}^6 \tag{3.2}$$

The three positional degrees of freedom (x, y, z) are often thought of and mapped separately from the three rotational degrees of freedom (ϕ, θ, ψ). Since the study describes the use of a 3DOF input device for controlling the 3D position of an object, it is intuitive to start by modifying Equation 3.2 to:

$$F = \begin{cases} f : \mathbb{R}^3 \rightarrow \mathbb{R}^3 & \text{where } f \text{ is the positional } (x, y, z) \text{ portion,} \\ g : \mathbb{R}^3 \rightarrow \mathbb{R}^3 & \text{and } g \text{ is the rotation } (\phi, \theta, \psi) \text{ portion.} \end{cases} \tag{3.3}$$

For each of the components in the framework (see Figure 3.1), the mathematical definition for the positional portion and the rotation portion is in essence the same. In the next section each component is defined mathematically and discussed in terms of the positional portion. However, the rotational portion can feel quite different to a person using it. Thus, when implementing an interaction technique that varies one or more of the components for the positional portion, one should not automatically apply the same to the rotational degrees of freedom. Therefore, when each

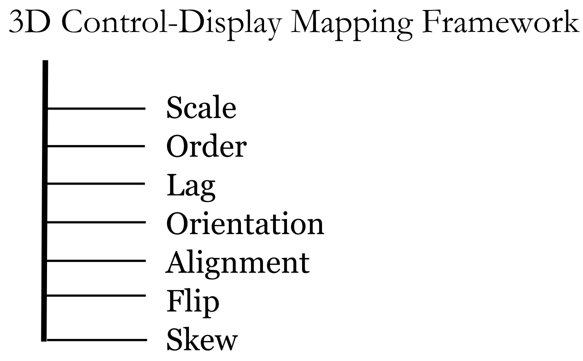


Figure 3.1: The 3D Control-Display Mapping Framework

component is discussed, if the rotational portion differs in a significant way from the positional portion it is noted in that section.

3.2 The Components of the Framework

Each of the components in the framework – scale, order, lag, orientation, alignment, flip, and skew – are discussed here. To illustrate how a component changes the 3D control-display mapping, this thesis introduces a new set of schematic diagrams.

Figure 3.2 shows an example of the schematic diagrams. This schematic diagram illustrates the standard mapping from interactive control space to virtual display space that occurs due to the perspective projection. The thick horizontal line represents the screen of a vertical display shown looking down on its top edge. In this example, the interactive control space is shown below the thick display line and the virtual 3D space is shown above the line. To illustrate the effect of perspective projection in the default control-display mapping dashed lines are used. In the default mapping the full interaction space (indicated by the dashed black lines) is mapped into the perspectively distorted box with blue dashed lines in the display space. The

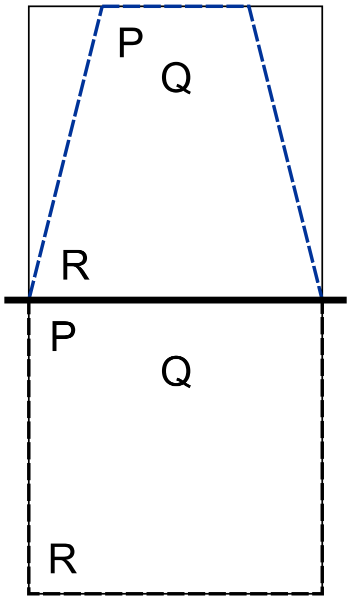


Figure 3.2: Schematic illustration of the standard default mapping. The thick horizontal line represents the display surface. Below the line is the interactive control space, above the line is the virtual display space. The thick dashed line represent the space users are moving in, both for control space (black) and the mapped interaction in display space (dark blue). The letters indicate positions of points in control space and the mapped positions of these points in display space.

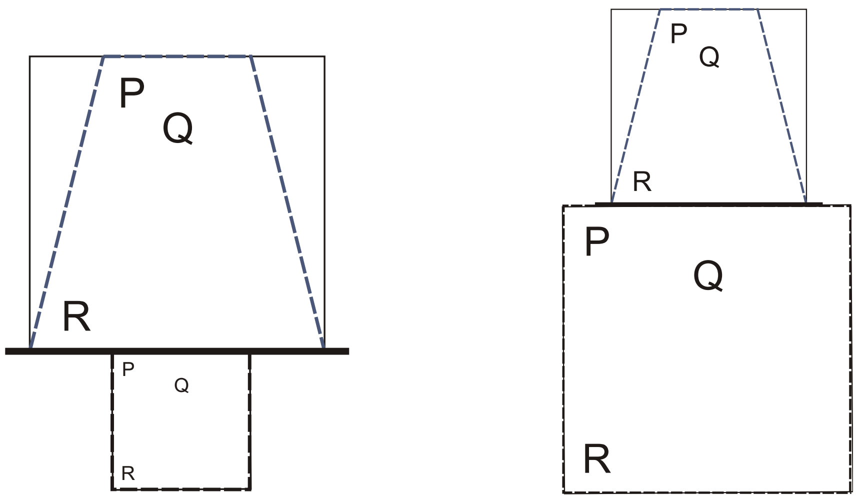
letters are used to indicate how specific points are mapped from control to display space. Notice that points in interactive space shift position and are closer together at large depth inside the perspective box in display space.

3.2.1 Scale

The control/display ratio [13], also known as control gain [54], defines the *scale* or size relationship between control and display space (See Figure 3.3). For example, linear scale functions are of the form:

$$f(\Delta x, \Delta y, \Delta z) = (s_x \Delta x, s_y \Delta y, s_z \Delta z,), \quad s_i \in \mathbb{R} \tag{3.4}$$

Although it is possible to have different scales for each axis, typically the same scale is used for all axes ($s_x = s_y = s_z$). The effect of scale on interaction, as well as its manipulation, is well understood and has been used previously in a number of systems. As discussed in section 2.4, examples include techniques such as World in Miniature (WIM) [76] which allows a user to interact at two different scales, the Go-Go technique [66], and other scaled metaphors which adjust scale to increase the range of direct manipulation [62, 58]. Wang and MacKenzie [82] studied the relative scale of input device, cursor, and target size and noted that a direct one-to-one scale between all three was easiest to use ($s_i = 1$).



(a) Small control space mapped to larger display space.

(b) Large control space mapped to smaller display space.

Figure 3.3: Schematic illustration of a small control space movement to a larger display space movement (like WIM) and vice versa. Notation as in Figure 3.2.

Unlike the positional *scale* portion, when rotational *scale* is introduced, it may be more common and quite beneficial for interaction to scale rotational DOF differently ($s_a \neq s_b \neq s_c$). For instance to overcome the rotational limitations of the human hand and arm, and allow a full 360 degrees, each rotational *scale* DOF would differ relative to the hand/arm axis and the range of physically possible values in that direction. The benefit and exact mapping would differ highly dependent on the input device. Poupyrev *et al.* [68] investigated the potential of rotational scaling. They note that scaling differs substantially between absolute and relative input. Absolute input maintains a nulling point, but does not preserve direction when scaled, and relative input does not maintain a nulling point but preserves direction when scaling [68]. They found that rotational scaling would be appropriate with relative input devices, particularly those that do not have an obvious neutral point, such as a spherical input device.

3.2.2 Order

It is also possible to differentiate between mapping the position in control space to the position, velocity, or acceleration in display space. Poulton [64] describes this difference as a change in *order* similar to the convention in physics and derivatives. For example, a velocity-based function (see Figure 3.4) is a first-order mapping of the following form:

$$f(\Delta x, \Delta y, \Delta z) = (\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t}, \frac{\Delta z}{\Delta t}) \tag{3.5}$$

A zero-order (position-based) control maps position in control space directly to position in virtual space. A first-order (velocity-based) control maps position in control space to the velocity of movement in display space. Lastly, a second-order (acceleration-based) control maps the position in control space to the acceleration of movement in display space. Zero-order control is the most common. First-order control is usually used for an isometric control such as a joystick. Second-order control is normally seen in racing games that replicate the accelerator on a car or other vehicle. Third-order, fourth-order, *etc.* mappings are mathematically valid and possible, and may be of use in particular application domains. It is possible to change rotational order (as mentioned by Poupyrev *et al.* [68]) but this is an unexplored direction.

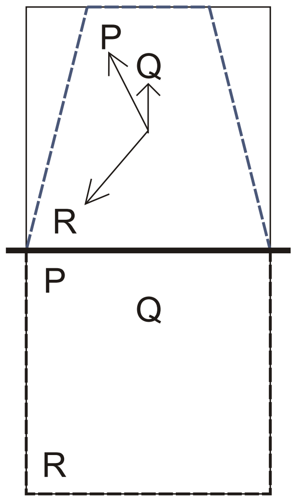


Figure 3.4: Schematic illustration of *order*. A first-order (velocity-based) mapping of control space to display space

As discussed in Chapter 2 (see Section 2.3), tracking position in control space is fairly difficult and current commercial tracking systems are expensive. Digital accelerometers on the other hand are relatively cheap, and thus there are 3D input devices that measure acceleration instead of position (*e. g.* the WiiMote). When using an input device, like the WiiMote, that provides acceleration this above discussion on order needs to be expanded and reconsidered. For instance, with an acceleration input device it is simple to use the acceleration data in control space and map it directly to acceleration in control space. If one wanted to try and use the acceleration data and obtain positional data or map to position in display space, one would have to integrate. Since integration commonly involves approximation and results in a set of solutions, this introduces positional uncertainty.

3.2.3 Lag

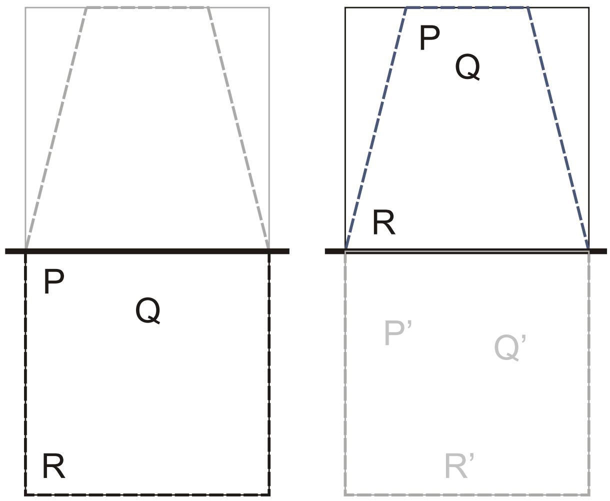


Figure 3.5: Schematic illustration of Lag. Motion in control space at an early time step on the left is displayed in display space at a later time step on the right

Lag can also occur either because of deficiencies in the input device or can be artificially introduced, on purpose, into the control-display space mapping. Such a function would require storage of the data obtained from the input device in the past (See Figure 3.5). One possible lag function could be as follows:

$$\begin{aligned} h_t(\Delta x, \Delta y, \Delta z) &= (\Delta x, \Delta y, \Delta z) \\ f_t(\Delta x, \Delta y, \Delta z) &= h_{t-L}(\Delta x, \Delta y, \Delta z), \exists L \in \mathit{positive}\ \mathbb{R} \end{aligned} \tag{3.6}$$

Lag is normally thought of as detrimental to interactivity, however in some cases, such as the presence of visual illusions (*e. g.* the Induced Roelofs Effect), it has been thought that the delay may actually improve performance [63].

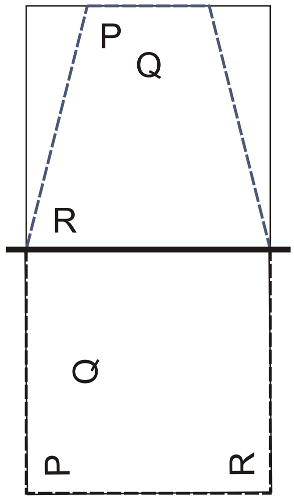


Figure 3.6: Schematic illustration of Orientation. A non-standard orientation. The control space is rotated about the height or y-axis

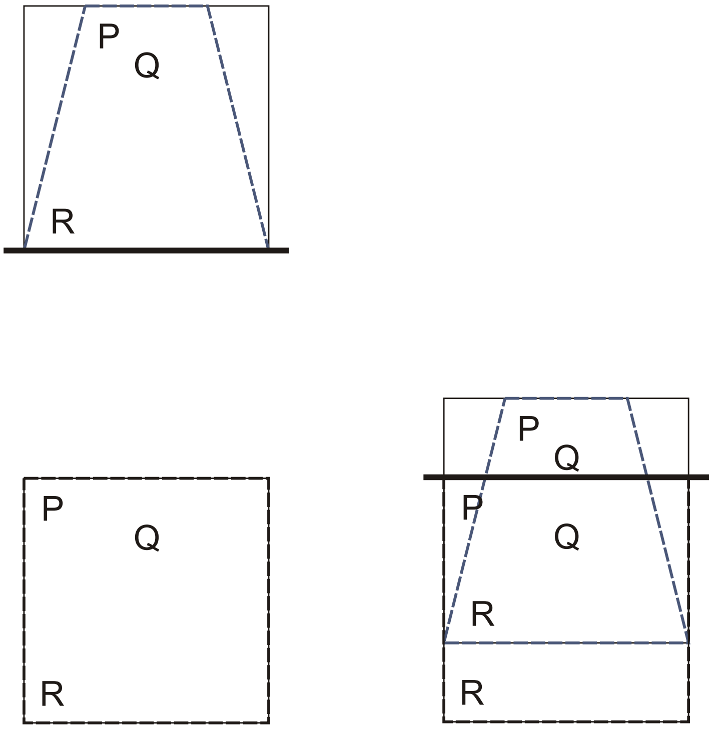
3.2.4 Orientation

Imagine two or more people are playing a 3D game and are passing the controller between each other while sitting at different orientations to the screen. The controller should be able to work in each person's orientation. *Orientation* refers to the orientation of the control space with respect to the display space. The entire space can be rotated by an angle around one of the coordinate axes (See Figure 3.6) or around an arbitrary axis. For example, a rotation about the z-axis by an angle of α would have the following control-display function:

$$f(\Delta x, \Delta y, \Delta z) = (\Delta x \cdot cos\alpha + \Delta y \cdot sin\alpha, \Delta y \cdot cos\alpha - \Delta x \cdot sin\alpha, \Delta z) \tag{3.7}$$

The effect of orientation has been studied in two dimensions by Wigdor *et al.* [88], but the 3D case has not been addressed to date. If the input device is 360 degree capable in all three rotational DOF then one can correct for the initial *positional orientation* by performing a rotation. However *positional orientation* effects the neutral rotation point, and a poor choice for *positional orientation* could result in the use of an awkward grip on the input device.

3.2.5 Alignment



(a) Control space and display space separated by some distance or aligned at a distance.

(b) Control space and display space overlap as in some HMD or Cave setups.

Figure 3.7: Schematic illustration of alignment. Notation as in Figure 3.2.

The distance between the interactive control space and the display space; or control-display space *alignment* could be defined as follows:

$$f(\Delta x, \Delta y, \Delta z) = \Delta \vec{f} + \Delta \vec{o} \tag{3.8}$$

where $\Delta \vec{f} = [\Delta x \ \Delta y \ \Delta z]$ and $\Delta \vec{o}$ is the vector $[(D_x - C_x) \ (D_y - C_y) \ (D_z - C_z)]$

where (D_x, D_y, D_z) is the origin of the display space and (C_x, C_y, C_z) is the origin of the control space. *Alignment* is usually restricted by the input and output display hardware and is often implicitly adjusted in the input calibration process. However, an alignment with a control space and display space that overlap, such as in a head-mounted display, would likely compliment different interaction techniques than an alignment where the interaction space is directly in front of a display or an alignment where the interaction space is at some distance from the display (See Figure 3.7). Rotational *alignment* is in essence positional orientation and as such is already described above.

3.2.6 Flip

The mapping can also *flip* one of the coordinate axes. Technically, this can be considered as a subset of scaling, but flipping an axis may affect the user's understanding in a profoundly different way than would scaling. For example, a flipping of the depth axis ($s_z = -1$) would cause the interaction to be like that in a mirror.

Rotational *flip* is also technically a subset of scaling (*e. g.* $s_a = -1$). However, in our everyday lives we have conventions about clockwise and counter-clockwise rotation. A variety of tools allow us to "flip" between the two. For instance, we do not need to think about the general 3D rotation of a ratchet wrench to know the tool is in clockwise mode. Thus, it may be appropriate to allow flipping to occur dynamically in some applications.

3.2.7 Skew

Finally, a *skew* mapping is a mapping that is rarely considered. For example, analogous to perspective projection of world coordinates to display coordinates, the mapping from control to display space may also be subjected to a perspective distortion, thus spatially skewing the mapping. The control-display mapping function would be:

$$f(\Delta x, \Delta y, \Delta z) = M \cdot \Delta \vec{x} \quad (3.9)$$

where M is the projection matrix and $\Delta \vec{x} = [\Delta x \ \Delta y \ \Delta z \ 1]_T$. In typical 3D applications, the control space is mapped using the same projection matrix as is used for mapping the world to the display. However, by applying an inverse distortion to the mapping we can control to what degree the projection distortion is present in the control-display mapping.

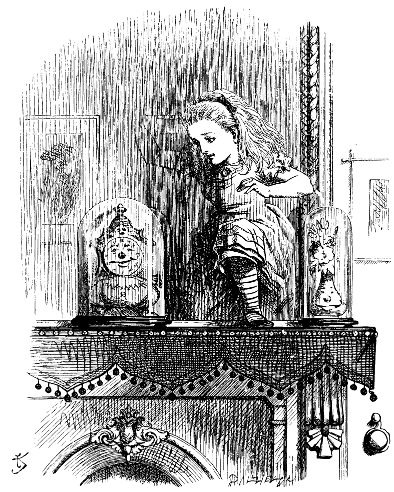
Standard perspective projection does not affect the rotational portion, and as such, the rotational portion normally would not be considered for variations in skew.

3.3 The Study's Four Selected Mappings

The last two of the positional mapping types (flip and skew) have not been discussed in the literature to date. This section will discuss the effects of these mappings in detail, and derive the four specific mappings examined in this study as special cases and combinations of these types.

When considering which mappings to explore in depth, metaphors from our everyday world that might lead to alternate but comprehensible control-display mappings have been sought. One choice is to address the issue shown in Figure 3.9 by

(a) Alice peering into the mirror.

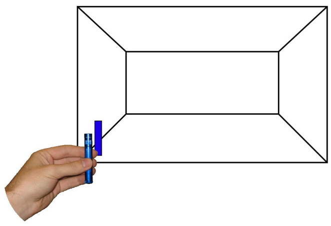


(b) Alice stepping through it.

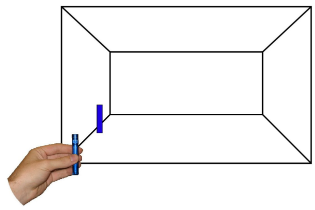
Figure 3.8: ‘Through the looking glass’ metaphor, from [24].

straightening out the distortion created by the common perspective mapping – an orthographic interaction in which the display is still in normal perspective but the control to display mapping is one-on-one for x- and y-movements. The second choice is to explore a ‘through the looking glass’ metaphor (Figure 3.8), which draws upon a classic piece of literature “Through the Looking Glass and What Alice Found There” [24] and from our experiences with mirrors in our everyday lives.

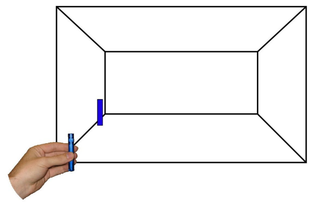
As explained above, the perspective projection often used as a visual effect by default in rendering also results in a distortion of the control-display mapping. This distortion causes forward and back motions in physical space to be mapped to perspective lines in virtual space. This agrees with the visual representation of the depth parameter. However, perspective lines move toward the center of the screen, and as such the cursor has a sideways motion on the screen while in control space the motion is perpendicular to the display surface. This disagrees with the kinesthetic representation which would require that a motion forward in real space is mapped



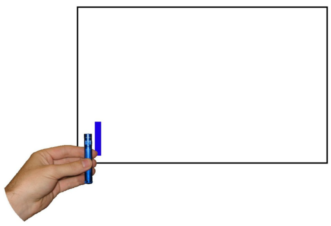
(a) Starting at the corner of the interactive space farthest from the screen, one is interacting near the front of the perspective box



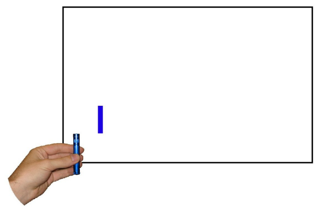
(b) Interacting somewhere in the middle



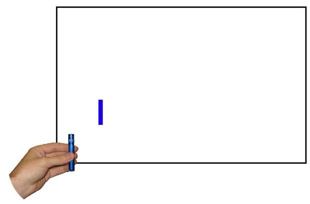
(c) Right up close to the screen, one is interacting at the back of the perspective box



(d) Starting at the corner of the interactive space farthest from the screen, the cursor appears directly in front



(e) Closer to the screen, the cursor begins to veer up and to the right



(f) Right up close to the screen, the cursor is quite far up and to the right

Figure 3.9: Mapping control space to display space: (a), (b), and (c) show a perspective display space with a perspective control/display mapping for the interaction. Note how the corresponding cursor object in display space makes sense within the box drawn in perspective but when the box cues are removed in (d), (e), and (f) this mapping is not so obvious. Note: This diagram is reproduced from Chapter 1 for convenience

to a motion directly forward or orthographically forward in virtual space. A control-display mapping that uses the perspective distortion originating from the perspective projection, will be said to be in *visual concordance*; and a controld-display mapping that uses an additional inverse perspective distortion to eliminate the effects of the perspective projection, will be said to be in *kinesthetic concordance*.

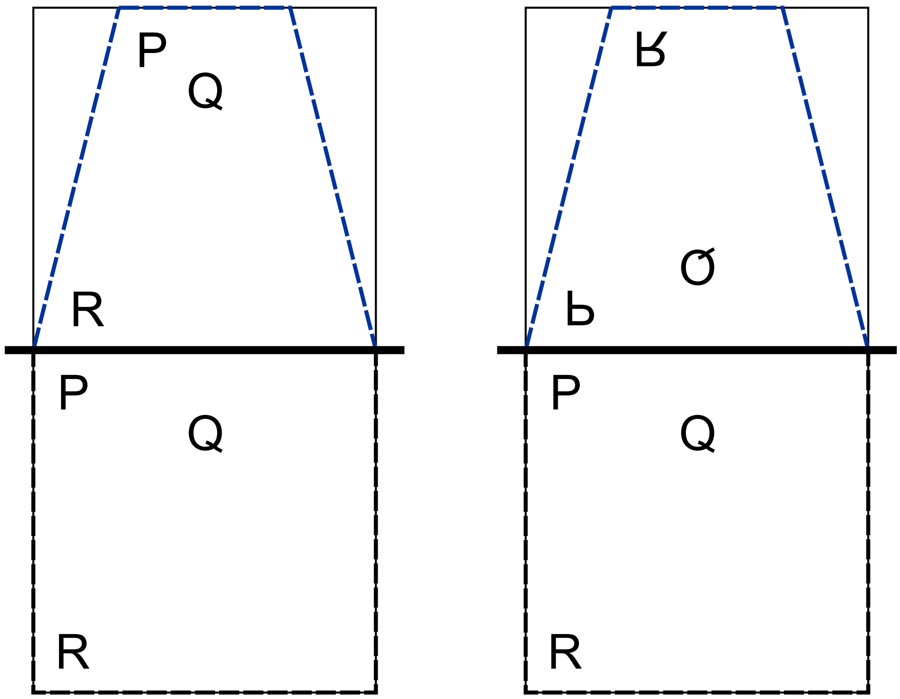
In addition to skew mappings, flip mappings,in particular, the flipping or mirroring of the depth dimension (z-axis) seems to be promising since we are used to interacting with mirrors in the real world. Thus, it seems natural to examine control-display mappings that mirror depth, such that a motion toward the screen in control space is mapped to a mirrored cursor motion toward the front of the scene in the display space. This mapping can then be compared to the more common mapping that translates a motion toward the screen in control space to a cursor motion away from the display surface. This difference is relevant for interaction with large displays where users are typically able to get close to the screen when interacting with objects.

The four mappings below are derived by combining the two flip conditions with the two skew conditions. Perspective mapping is the de facto standard mapping. Mirrored perspective mapping is the mirrored version of this standard mapping, where the depth axis is mirrored. Both of these agree with visual concordance. Orthographic mapping and mirrored orthographic mapping are the result of eliminating the mapping distortion caused by perspective projection and its mirrored version, both of which represent kinesthetic concordance.

These alternative control-display mappings have not been implemented before. For the default mapping (*i. e.* perspective mapping) one can use the built-in rendering system as is. As each mapping is described in the next sections, the implementation details necessary to implement each mapping is briefly noted.

3.3.1 Perspective and Mirrored Perspective Mapping

Figure 3.10(a) shows a schematic view of the standard mapping from interactive control space to virtual display space that occurs due to the perspective projection. The letters are used to indicate how specific points are mapped from control to



(a) Perspective skew mapping/visual concordance.

(b) Mirrored perspective skew mapping/mirrored visual concordance.

Figure 3.10: Schematic illustration of the perspective and mirrored perspective mapping.

display space due to skew and flip differences between the mappings. Notice that points in interactive space shift position and are closer together at large depth inside the perspective box.

Figure 3.10(b) shows the standard mapping mirrored in the z-direction. Notice that the points in interactive space are still all inside the perspective box, but that symbols and positions are now mirrored. A point near the screen in real space is now close to the screen in virtual space.

Implementation Notes

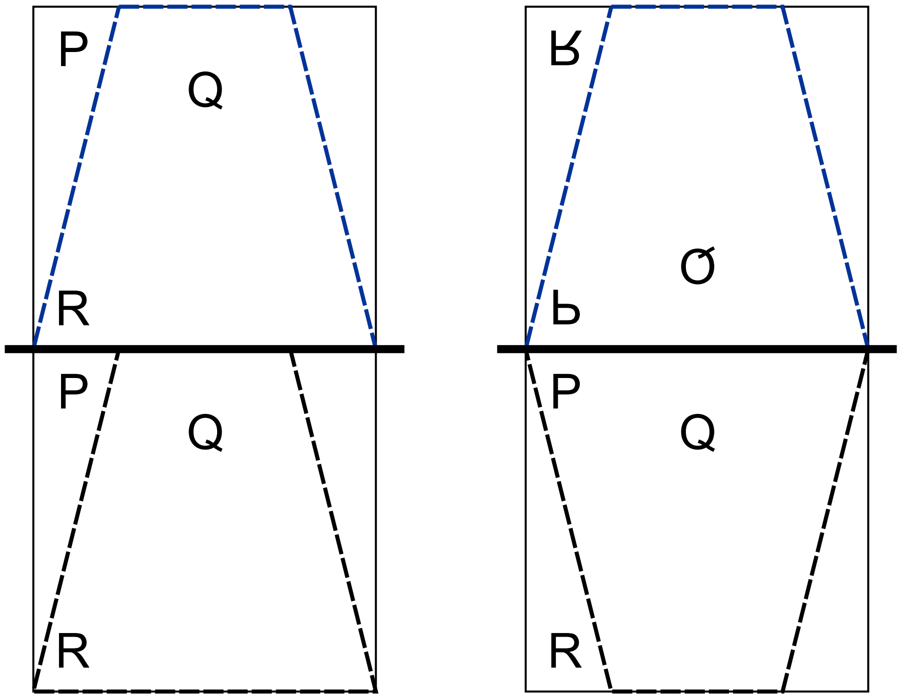
High-end tracking systems have their own calibration system that allows one to map physical real world coordinates to match the virtual coordinates in supported graphics systems (*e. g.* OpenGL, DirectX, Java3D, *etc.*). If one implements a tracking system, as was done for this thesis, careful calibration is necessary to convert tracking system coordinates to graphics system coordinates. In the calibration process, a choice for the positional *alignment* between control and display space is set. Although, if desired, the *alignment* could be modified programmatically at any time. In the thesis implementation the control and display space are aligned so they meet at the screen. To modify the 3D control-display mapping, the dimensions of the control volume and the display volume must be known. Further, one needs to know the location of each volume in the respective coordinate system.

Once this calibration has occurred, the perspective mapping is constructed using the built-in rendering system. For the mirrored perspective mapping one needs to flip a point in the depth axis. To do this one needs the depth coordinate of the point in control space (p_z), the depth coordinate of the near side of the display volume (n_z), and the length of the display volume's depth dimension (v_z). Let m_z be the new mirrored point, then:

$$m_z = n_z + v_z - (p_z - n_z) = 2n_z + v_z - p_z \tag{3.10}$$

3.3.2 Orthographic and Mirrored Orthographic Mapping

Figure 3.11(a) demonstrates that, through an additional inverse perspective distortion the effect of the perspective projection is removed, resulting in an undistorted mapping from control to display space. The points in interactive space are now mapped directly or orthographically into virtual space (note: everything in the display is still drawn in perspective projection, this inverse perspective distortion only



(a) Orthographic skew mapping/kinesthetic concordance.

(b) Mirrored orthographic skew mapping/mirrored kinesthetic concordance.

Figure 3.11: Schematic illustration of the orthographic and mirrored orthographic mapping. Notation as in Figure 3.2.

effects the resulting location of interaction mapped into the virtual display space). Notice that perpendicular motions in control space (from R to P) now also map to perpendicular motions to the display surface in display space, which matches the kinesthetic model. This also has the effect that the virtual cursor is always directly in front of the physical cursor. Finally notice that the perspective reference box in virtual space now also maps back to a perspective box in interactive space.

In Figure 3.11(b) an additional mirroring is applied to the z-axis, resulting in a mirrored orthographic mapping. Notice that motion paths perpendicular to the display surface in control space still map to motion paths perpendicular to the display surface in display space. However, they are mirrored at the screen surface such that points close to the screen in interactive space are close to the screen in display space.

Implementation Notes

The orthographic mapping applies an inverse perspective distortion to cancel the effect of the perspective projection. If one knows the skew matrix M, one can calculate M^{-1}. Then simply before rendering, one multiplies the point by this inverse matrix. Alternatively, as was done in this thesis, one can use gluUnProject() in OpenGL, which is a function that takes the physical 2D screen coordinates, and a z-buffer coordinate and recovers the virtual world coordinates. To use this function, one must determine these screen coordinates x_{screen}, y_{screen} and the z-buffer coordinate z_{buffer}, and then one applies gluUnProject() before rendering. To determine the screen coordinates, points are first projected using gluProject(). One can then calculate the screen coordinates using the newly projected coordinates x, y, z, the known coordinates of the screen edges $x_{leftedge}, y_{bottomedge}$, the display volume width and height $w_{display}, h_{display}$, and the screen width and height w_{screen}, h_{screen}.

$$x_{screen} = ((x - x_{leftedge})/w_{display}) * w_{screen} \tag{3.11}$$

$$y_{screen} = ((y - y_{bottomedge})/h_{display}) * h_{screen} \tag{3.12}$$

To calculate the z-buffer coordinate the following z-buffer formula is used:

$$z_{buffer} = (1 - near/(-z))/(1 - near/far) \tag{3.13}$$

For the mirrored orthographic mapping, one first applies axis flipping, then one applies the inverse perspective distortion.

Chapter 4

Study

In this chapter, the evaluative study that was conducted to investigate some of the unexplored possibilities suggested in the control-display mapping framework described in Chapter 3 is discussed. This chapter begins with the motivation to conduct the study, followed by the study design. It continues by describing the experiment: the participants, the apparatus, the procedure, and data collection. The study results are then presented and the chapter concludes with a discussion of the findings.

4.1 Motivation

In this study of possible alternative 3D control-display mappings, the effects of modifying the *flip* and *skew* components in the 3D control-display mapping are investigated. The reasons are as follows:

1) Although 3D interaction is a mature field, people still have difficulty interacting in computer-mediated 3D.

2) The control space is commonly mapped directly into a perspectively distorted display space without much thought on the effect this distortion may have on interacting in 3D.

3) Previously, only variations in scale have been explored. The effect of flipping axes have not been investigated. Likewise, other than the default perspective

skew that results from mapping directly into a perspectively distorted display space, the effect of other skews have not been investigated.

4) To see if familiarity with mirrors translates into effective 3D interaction.

5) To see if decoupling the visual perspective distortion from the interactive mapping produces intuitive 3D interaction.

4.2 Study Design

Although the 3D control-display mapping framework applies to 3D interaction in general, this study specifically looks at selection in the *within reach* metaphor (as defined in Section 2.4.4). The tasks (see section 4.5) reflect this and are based upon the most common study selection tasks, namely a point or volume location target [16, 17, 39]. A point selection task involves simply selecting a point by placing a cursor at that point. A volume is the same, except one places the cursor within the volume. Fitt's Law is a well accepted model of point or area selection, has been studied extensively in 2D, and extended to 3D [39]. In the tradition of Fitt's Law studies, two of the tasks are based on the 2D multi-directional tapping task described in the International Organization for Standardization's (ISO) Standard 9241 Part 9 [74] (See Figure 4.1).

The goal of this study is to investigate people's performance as a result of differences in skew and flip, so all other components described in the 3D control-display mapping framework remain constant. Since mirrored mapping is of interest, in particular, when users interact close to the surface of the screen, the distance (alignment) between control and display space is kept to zero (*i. e.* the control and display touch

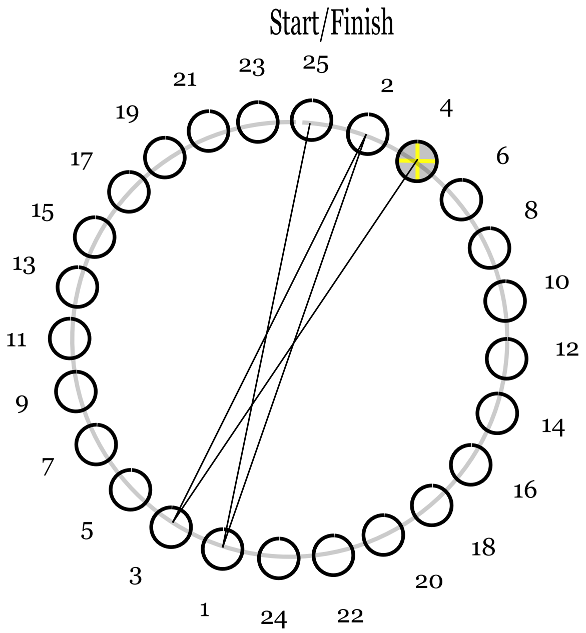


Figure 4.1: The 2D multi-directional tapping task described in ISO Standard 9241 Part 9 [74]. In this task, participants tap a series of circular targets which are placed on a larger circle. A participant starts at the circle at the top labeled 25, and then taps 1, and then 2, and so on, crisscrossing over the whole large circle. In the experiment targets are not numbered, instead the next target is highlighted with cross and potentially a different colour to clearly indicate which target is next. The idea is that the experiment controls for direction. The 3D variations used in this study add a variable height, or rising pillars to create the third dimension.

each other, but do not overlap). The lag component is not artificially increased and any minimal lag present is due to the tracking system. A scale factor of 1 for all axes and a position-based control are used.

A 2 (*flip:* mirrored, non-mirrored) $\times$ 2 (*skew:* perspective, orthographic) within-subjects design was used resulting in a total of four mappings.

4.3 Participants

Twenty-four people (13 male, 11 female) participated in the study, recruited through e-mail, the departmental "notice of the day", and word of mouth. Four of the participants were left-handed and the remaining 20 were right-handed. All were experienced computer users. Six used computers 5-15 hours a week and six used computers 15-35 hours a week while the other twelve used computers more than 35 hours a week. Fifteen participants reported having little to no previous 3D gaming/graphics experience (novice users) and nine reported having moderate to a lot of previous 3D gaming/graphics experience (experienced users). The participants included first year undergraduate computer science students, graduate computer science students, computer science alumni, new media artists, and computer gamers.

4.4 Apparatus

The display used for all four tasks was a 1024×768 pixels, $73.3\,\text{cm} \times 55.0\,\text{cm}$ wall display. In all cases, the graphics that created the virtual 3D display space formed a perspective grid representing five walls of a virtual room with lighting as an additional depth cue (see Figure 4.2). Participants were able to control a 3D cursor using a tracked light pen in the $73.3\,\text{cm} \times 55.0\,\text{cm} \times 55.0\,\text{cm}$ volume directly in front of the display. Participants were given the option to sit, but all chose to stand throughout all trials. The pen was tracked using vision algorithms and input from two cameras mounted directly above the control space. Infrared filters were used to detect the near infrared light from the pen light (non-LED). Two computers were used in the experiment: one to capture movement via the two cameras and transmit the captured movement data to a second (main) computer, which controlled the display environment. According to the four studied mappings (non-mirrored perspective, mirrored perspective, non-mirrored orthographic, and mirrored orthographic), movement in the control space was mapped to the movement of a cursor in the display space. The cursor was represented at a one-to-one scale (as suggested by Wang *et al.* [82]) in the display space as a cylinder texture-mapped with an image of the light pen. Crosshairs have been used previously in 3D selection studies [39]. In this study crosshairs extending to each wall of the display space are used as additional feedback for the position of the cursor. The input used was an absolute indirect untethered tracking device.

4.5 Procedure

Before beginning to interact with the system, participants were asked to complete a short questionnaire to collect some demographics. For each mapping, participants performed four tasks in the same order, followed by a break task (see Table 4.1). After the break task, the participant was asked to fill out a questionnaire to garner feedback about the particular mapping they had just used. After all tasks were completed in all four mappings a participant was asked to complete an overall questionnaire to garner commentary and comparisons of the mappings. The order of mappings was different for every participant, so that all 24 possible orderings were each used exactly once. This exhaustive set is a control for order. At the beginning of the experiment the experimenter demonstrated how to interact with the system and suggested how a person could hold the input device. However, the experimenter did not specify that

Step	Mapping	Task
1	1	Grid
2	1	Static ISO
3	1	Moving ISO
4	1	Photo Box
5	Break Task	
6	2	Grid
7	2	Static ISO
8	2	Moving ISO
9	2	Photo Box
10	Break Task	
11	3	Grid
12	3	Static ISO
13	3	Moving ISO
14	3	Photo Box
15	Break Task	
16	4	Grid
17	4	Static ISO
18	4	Moving ISO
19	4	Photo Box
20	Break Task	

Table 4.1: The order in which tasks were done by participants.

a person must hold it a certain way, nor did the experimenter correct participants. Before each mapping was used, the experimenter demonstrated the mapping to the participant.

First Task—Grid Task: For the Grid Task each participant was asked to tap a sequence of 13 orange squares on the checkerboard walls with the bottom[1] of the light cursor. A single orange square would appear at a time in a grid section of the alternating gray grid on the walls of the room (see Figure 4.2). An orange square could be tapped by placing the cursor within an invisible rectangular volume that extended slightly below and above the surface of the orange square.

Second Task—Static ISO Standard Task: Inspired by the multi-directional

[1]The choice between whether to use the top or bottom of the light cursor was based on the fact that using the bottom allowed for better tracker accuracy.

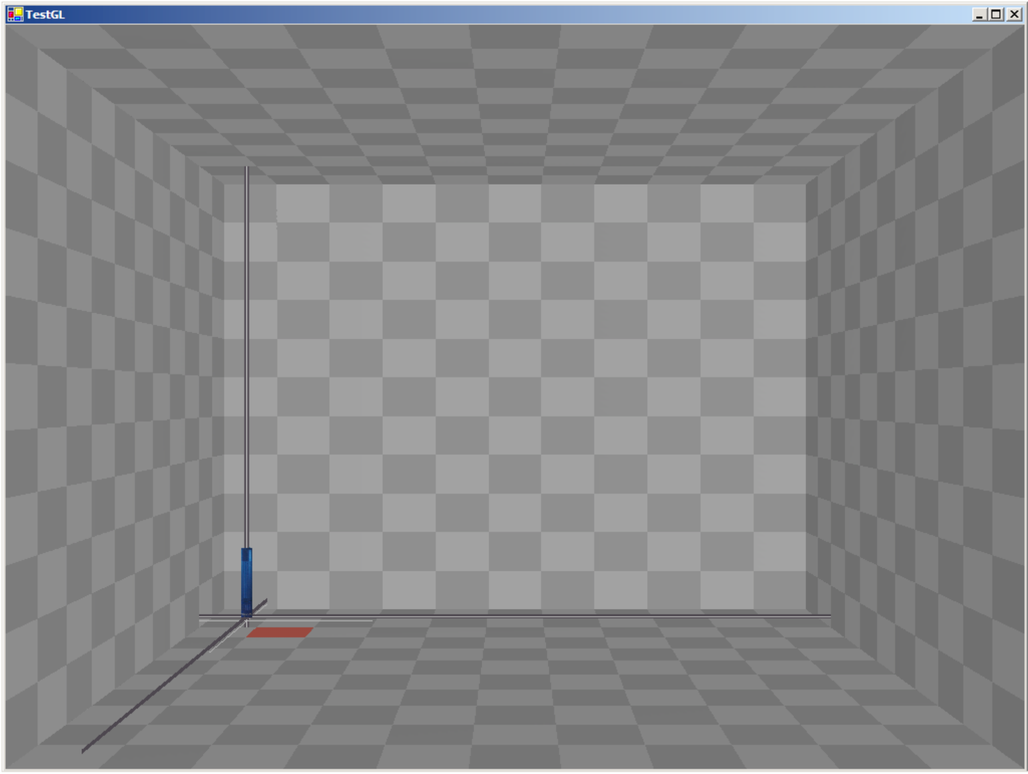


Figure 4.2: Grid Task. Participants were asked to tap orange squares that appeared on the checkerboard walls with the bottom of the blue light cursor.

tapping task described in the ISO Standard 9241 Part 9 [74], participants were asked to tap the top center of cylinders placed on a circle. To indicate which cylinder was next, the cylinder would pop up to a random height with a crosshair on the top (see Figure 4.3). To successfully tap the top of the cylinder, the cursor point had to be within a minimal spherical distance from the center of the crosshair on the cylinder.

Third Task—Moving ISO Standard Task: This task is similar to the previous task, except that instead of a pillar popping up to a random height, it would rise slowly from the floor, creating a moving target.

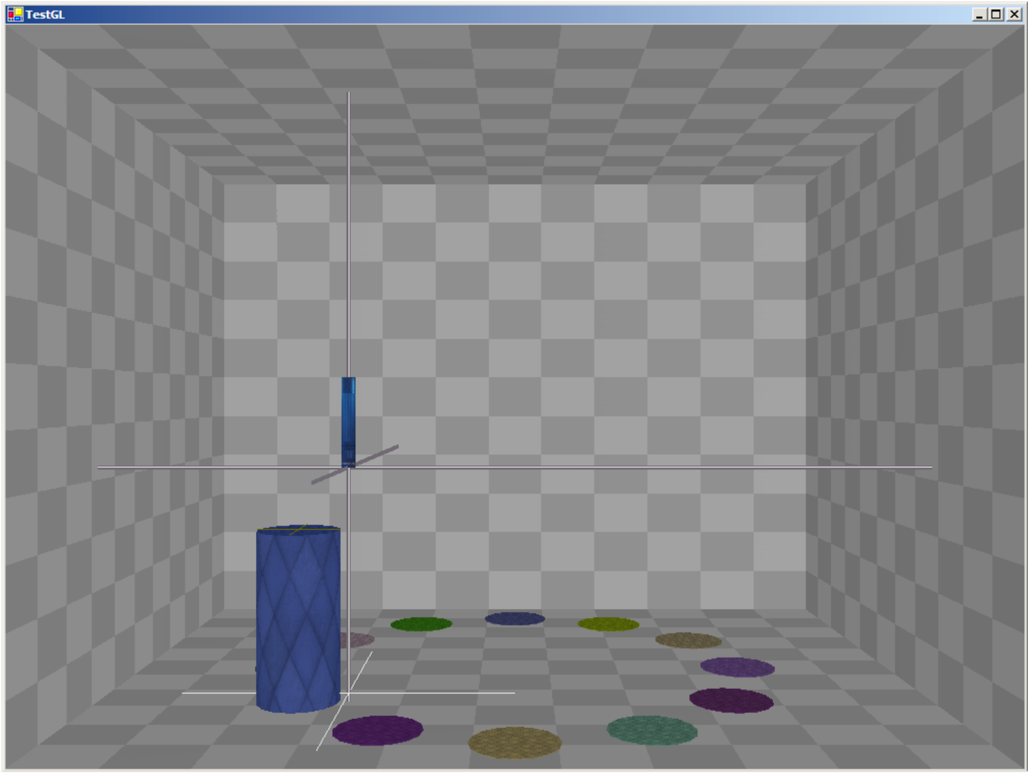


Figure 4.3: Static or Moving ISO Task. Participants were asked to tap the top of the raised cylinder with the bottom of the light cursor. The cylinders that were to be tapped were either not moving (Static ISO Task) or slowly rising from the ground (Moving ISO Task).

Fourth Task—Photo Box Task: In this task a set of image cubes were randomly placed on the floor (bottom) of the virtual room. As a slight change from the previous tasks, participants were asked to sort the image cubes into whether they thought the image cube had an inorganic or an organic image upon it (see Figure 4.4). They would do this by picking up the cube and placing it either on a blue mat or green mat at the front of the virtual room.

Break Task—Sketch Task: This task was intended as a break between sets of tasks, in order to rest the arm of the participant, as well as to garner feedback on

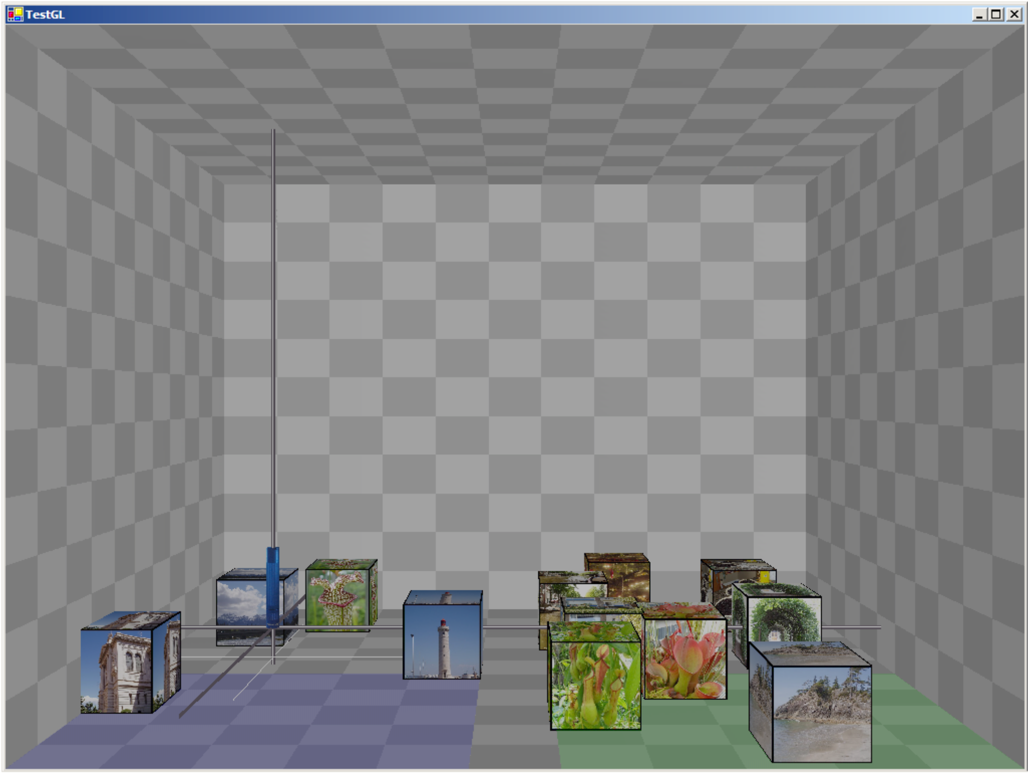


Figure 4.4: Photo box task. Participants were asked to sort the textured cubes into those showing inorganic or organic images and according to this decision place them on either of the two colored mats at the front of the virtual room.

the schematic control-display diagrams. To garner this feedback, the experimenter drew out a sketch of the schematic control-display and explained each part as it was drawn: the interaction (control) space, the screen, the virtual room, the space behind the walls, and two points in interaction (control) space. Participants were then asked to draw the location of two points in the the virtual (display) space, that correspond to the two indicated points in interaction (control) space.

4.6 Data Collection

Each session was video recorded to capture the motion of the user, but also to capture audio comments from the participants as they were using the system. The system logged the virtual tracker motion, as well as any events that occurred (*e. g.*, picking up a box or tapping a cylinder). Questionnaires were also used before the interaction began, after each break task, and at the end of the study to gather demographics, preference statistics, and open comments. Thus, our assessment is based on both quantitative data (with the caveat of the limitations discussed in Section 4.7) and qualitative data collected during the experiments and afterward with the questionnaires.

In the grid task, static ISO task, and the moving ISO Task, the small volume around the target used as the selection mechanism sets a specific accuracy. Thus the completion times used are a combined speed and accuracy measure. This is similar to the combined speed and accuracy measure, throughput, discussed alongside the ISO standard test [53, 74].

4.7 Results

An exploratory study is undertaken with a small sample of 24 participants drawn primarily from computer science students and alumni. Definitive conclusion cannot be made from this study because it is based on a small non-random sample, and differences seen in this sample may not be statistically significant. The quantitative results reported here provide useful insights and a larger study might substantiate the trends and findings found in this section but it is impossible to make conclusive statistical findings based on our small non-random study. However, it is important to note that the results are interesting and provide substantial evidence that further, more detailed study and analysis is merited.

A 2 (*flip:* mirrored, non-mirrored) × 2 (*skew:* perspective, orthographic) within-subjects factorial analysis of variance (ANOVA) [79] was performed on the task completion times for the grid task and both ISO tasks. Previous 3D interaction studies with a small sample size [29, 32, 75, 78], have defined "significance" to be at a p value less than 0.05 while values above this but less than 0.10 are considered marginal. We will adopt this terminology here when appropriate but it is important to note that these values are arbitrary and will need to be validated with a future study.

4.7.1 Grid Task

There was a significant mirrored versus non-mirrored effect in the grid task ($F_{1,23} = 13.1$, $p = .001$) and a marginal effect of orthographic versus perspective ($F_{1,23} = 3.3$, $p = .08$). The interaction was not significant ($F_{1,23} = 1.1$, $p = .30$). Participants performed the task more quickly when the mapping was not mirrored ($M = 8.57s$, $SD = 3.28s$) than when it was ($M = 10.69s$, $SD = 4.03$ s). Participants performed the grid task more quickly in the orthographic mappings ($M = 9.18s$, $SD = 3.70s$) than in the perspective mappings ($M = 10.07s$, $SD = 3.90s$). Figure 4.5 and Table 4.2 show the mean task completion times for each mapping.

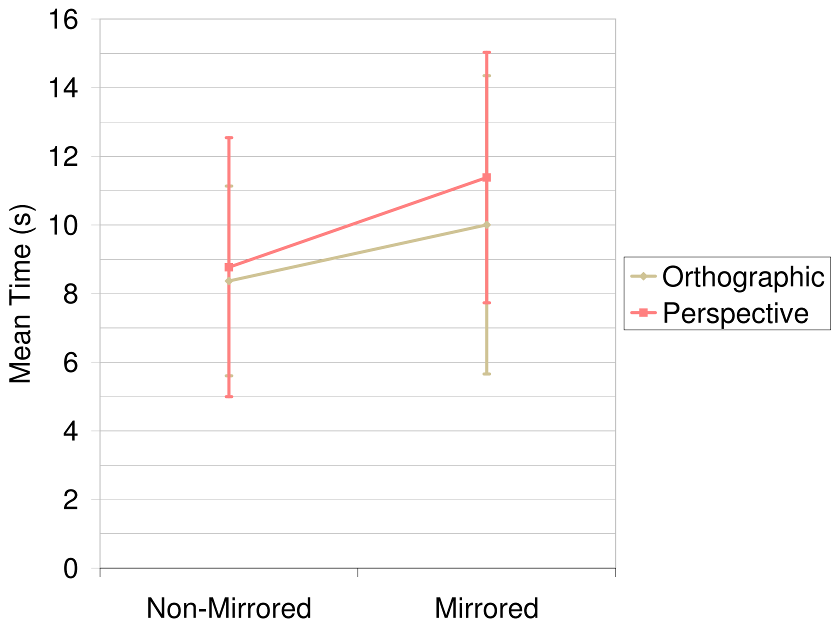


Figure 4.5: Grid Task: mean task completion times for the grid task. The mirrored mappings were slower than the non-mirrored and the orthographic mappings were faster than the perspective mappings. There was no interaction between mirrored versus non-mirrored; and orthographic versus perspective.

4.7.2 Static ISO Task

For the static ISO task, there was a noteworthy mirrored versus non-mirrored effect ($F_{1,23} = 5.5$, $p = .03$). There also were differences between orthographic versus perspective effect ($F_{1,23} = 6.0$, $p = .02$) and some distinction in the interaction ($F_{1,23} = 3.4$, $p = .08$). Post-hoc analysis revealed that, in the non-mirrored case, the difference between orthographic and perspective was not a factor ($p = .68$), but for the mirrored case, participants performed faster with the perspective than with the orthographic ($p = .01$). Figure 4.6 and Table 4.2 show the mean task completion times for each mapping.

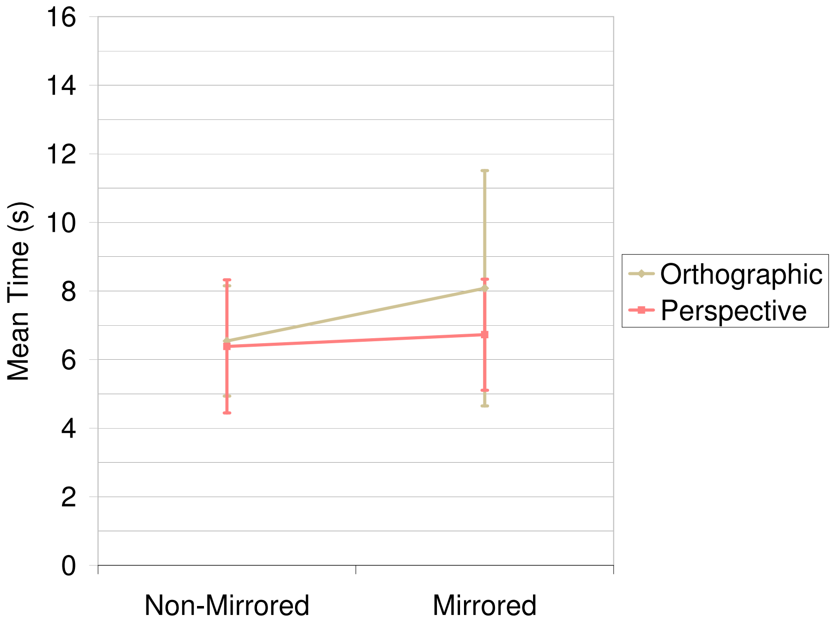


Figure 4.6: Static ISO Task: mean task completion times for the static ISO task. The mirrored orthographic mapping was slower than the non-mirrored orthographic, non-mirrored perspective, and the mirrored perspective mappings.

4.7.3 Moving ISO Task

There were notable differences between mirrored and non-mirrored effects in the moving ISO task ($F_{1,23} = 15.5$, $p = 0.001$). Participants completed the task more quickly when using a non-mirrored mapping ($M = 6.28s$, $SD = 1.58s$) than when using a mirrored mapping ($M = 7.96s$, $SD = 2.72s$). There was no orthographic versus perspective effect ($F_{1,23} = 0.1$, $p = .76$) and no interaction ($F_{1,23} = 0.4$, $p = .54$). Figure 4.7 and Table 4.2 show the mean task completion times for each mapping.

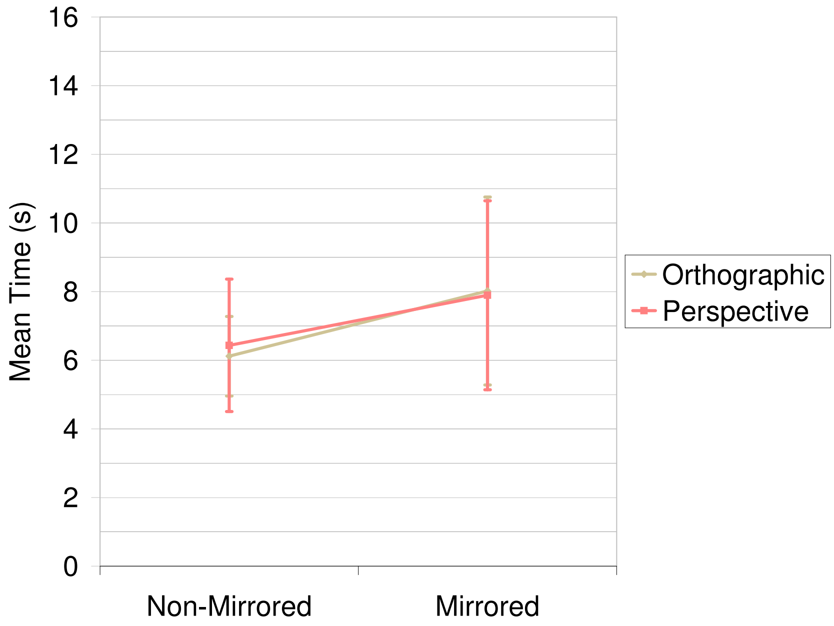


Figure 4.7: Moving ISO Task: mean task completion times for the moving ISO task. The mirrored mappings were slower than the non-mirrored mappings, and there was no orthographic versus perspective effect.

Outcome Measure		**Non-Mirrored(NM)**			**Mirrored(M)**			**M vs. NM Effect**			**O vs. P Effect**			**Interaction**		
		Mean(s)	SD(s)	n	Mean(s)	SD(s)	n	F	(df)	P*	F	(df)	P*	F	(df)	P*
Grid	Orthographic (O)	8.37	2.76	24	10.00	4.34	24	13.1	(1,23)	.001	3.3	(1,23)	.08	1.1	(1,23)	.30
	Perspective (P)	8.77	3.77	24	11.38	3.65	24									
Static ISO	Orthographic (O)	6.54	1.61	24	8.08	3.43	24	5.5	(1,23)	.03	6.0	(1,23)	.02	3.4	(1,23)	.08
	Perspective (P)	6.38	1.94	24	6.72	1.62	24									
Moving ISO	Orthographic (O)	6.12	1.16	24	8.02	2.74	24	15.5	(1,23)	.001	0.1	(1,23)	.76	0.4	(1,23)	.54
	Perspective (P)	6.43	1.93	24	7.89	2.75	24									
	* significance level at .05 level															

Table 4.2: Outcome measures by task (ANOVA)

4.7.4 Photo Box Task

The Photo Box Task was a simple manipulation task where the participant had to think about something different instead of simply selecting and moving to the next target. Participants chose which photobox to pick up first, whether to pick up a photobox at all, and whether to pick up a photobox multiple times. Thus, mean completion time is not an appropriate statistic to measure. Instead, this open ended task was intended to elicit more comments and to check whether participants' strategies changed when not explicitly thinking about selection of targets. Comments collected during this time are discussed in the quotes section below, and a participant's strategy did not appear to differ from the other tasks. Participants that did well in a particular mapping in the other tasks did well in the mapping in the photo box task. Participants who were having difficulty with a mapping in the other tasks, similarly, also had difficulty in the photo box task and used consistent strategy throughout the tasks.

4.7.5 Subjective Preference

After completing all the tasks in all four mappings participants were asked to rank the four mappings in order of preference (see Figure 4.8). Participants showed a clear preference for the non-mirrored mappings (92% first, 83% second, 8% third, 17% fourth). Most participants chose the non-mirrored orthographic mapping as their first choice (58%) followed by the non-mirrored perspective (33%) and the two mirrored mappings were each chosen as first by only one participant.

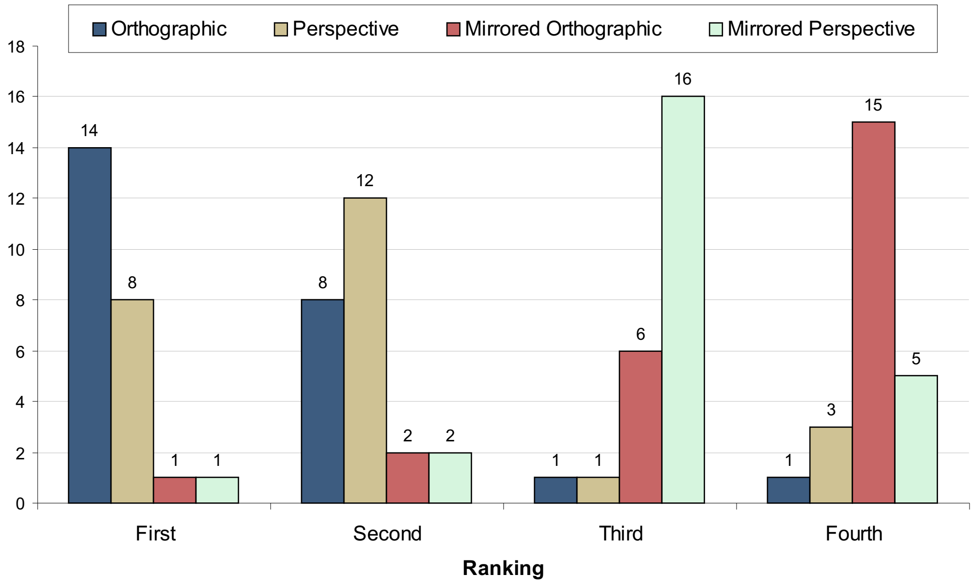


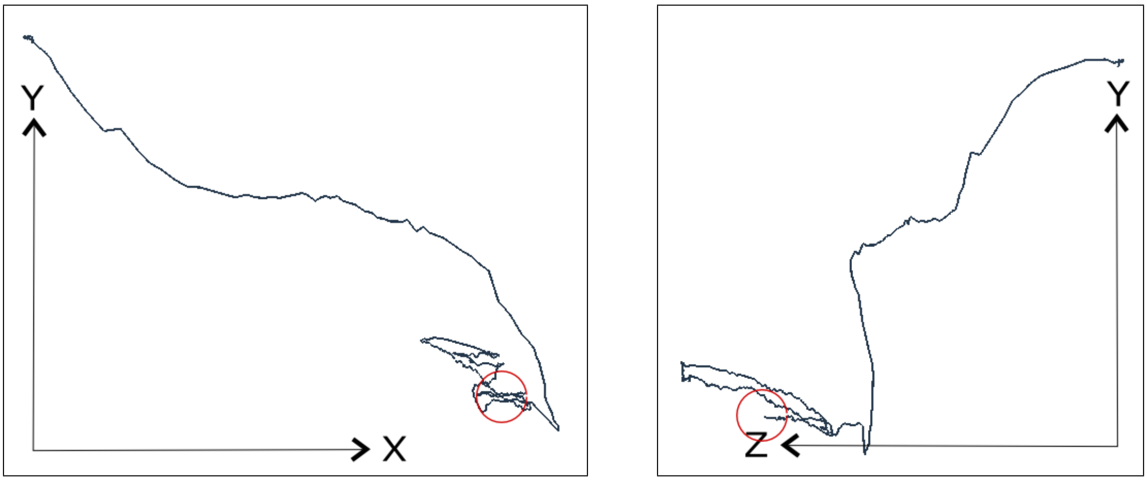
Figure 4.8: Overall mapping preference ranking: participants were asked to order the mappings from most preferred (first) to least preferred (last).

4.7.6 Trajectory Traces

From the logged data it is possible to reconstruct and trace the path that participants took between target points. Typically these trajectories show that participants generally had little difficulty going from one target to the next and took an almost direct path to get there. However, in every mapping, almost all participants would at some point have difficulty tapping a target. Some participants consistently were more direct in both non-mirrored mappings, others in both orthographic mappings, and still others in both perspective mappings. However, even in the mappings a particular participant was most direct in, he/she would have difficulties which seemed to appear at random times and for all tasks.

In each of the trajectory trace Figures (4.9 to 4.12), based on example data of difficulty in each mapping, two orthogonally rendered views are shown, one from the front and one from the side of the trajectory a participant took between two targets. The orthogonal rendering allows us to look at x-y-projection in the front view and y-z-projection in the side view. The target is circled in red.

In Figure 4.9 an example of difficulty in the non-mirrored orthogonal mapping is shown. The participant quickly and directly goes to approximately the right location and then has difficulty in the depth dimension z and some difficulty in the horizontal dimension x.

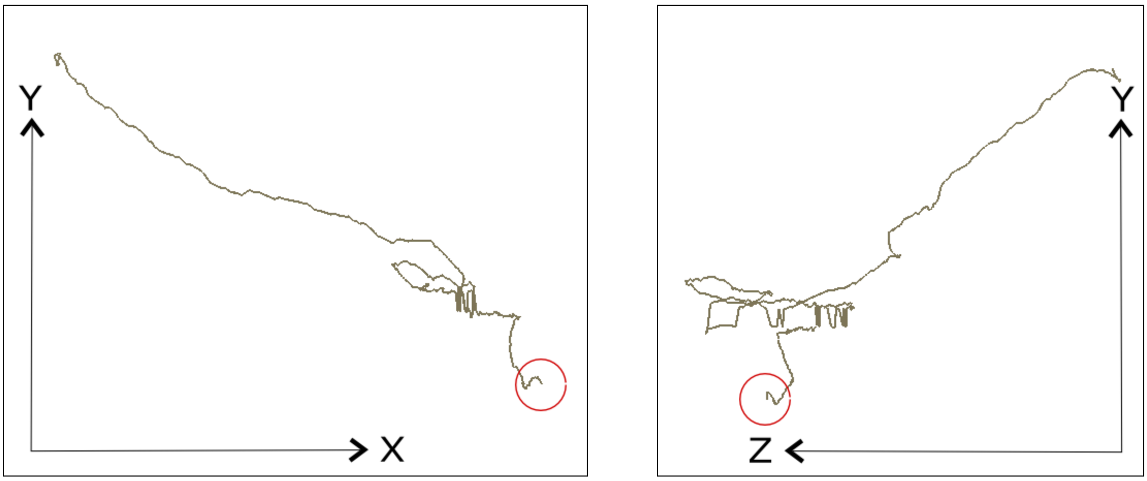


(a) Front view. (b) Side view.

Figure 4.9: Trace between two targets in orthographic mapping. Notice the amount of motion around the target, particularly in the z-direction.

In the example trajectory in Figure 4.10 of the non-mirrored perspective mapping, a participant gets close to the target and then has a lot of movement in z and y just above the target before they can orientate themselves again and move directly down to the target.

In the mirrored cases Figures 4.11 and 4.12, from the front view it appears the participant goes almost directly there with some adjustment in y, but in the side



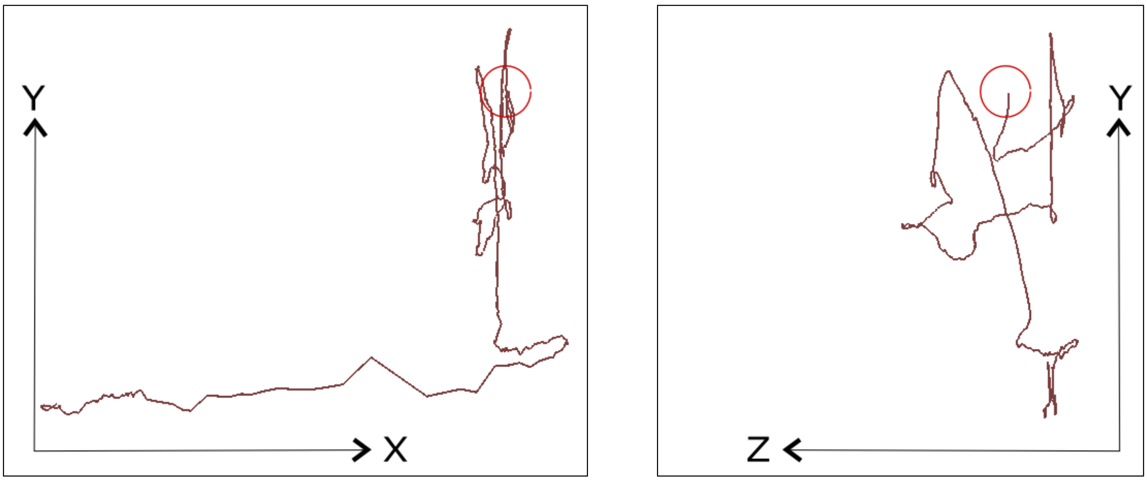
(a) Front view. (b) Side view.

Figure 4.10: Trace between two targets in perspective mapping. Notice that the user becomes lost in depth slightly above the target.

views one can notice that the participant is in fact having considerable difficulty with the depth dimension z.

4.7.7 Quotes

Participants had a wide variety of responses to the differences in mappings. Generally as reflected in the ratings above (see Section 4.7.5), mirrored mappings gave participants the most difficulty: "The A and B [non-mirrored] seemed more user-friendly, and the C and D [mirrored] seemed counteractive to my visual way of spatial sorting." Some participants therefore strongly disliked it, "Damned mirrored!," others enjoyed this challenge, "kinda fun because challenging" (mirrored perspective), "I liked the other [non-mirrored] method better but this one [mirrored] I got more of a sense of accomplishment once I grasped it!" However, a few users in fact did better in and preferred the mirrored mappings. A novice user wrote "[I am] used to mirror image interaction, because [I am] used to seeing objects move away when moving

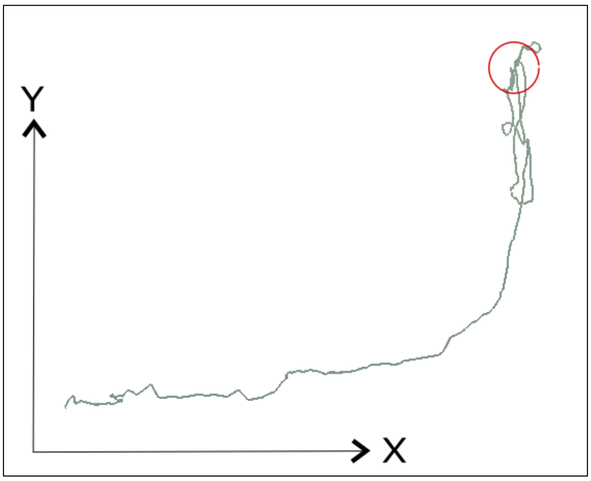


(a) Front view. (b) Side view.

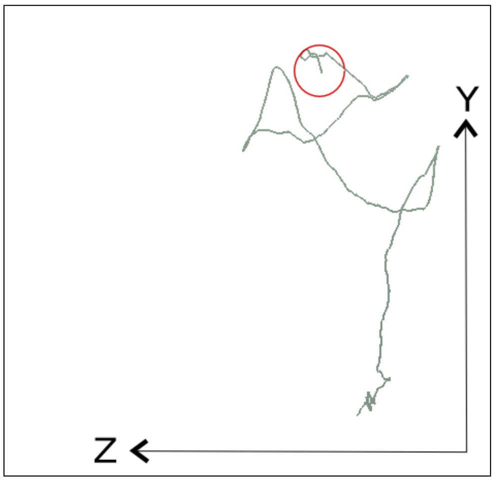
Figure 4.11: Trace between two targets in mirrored orthographic mapping. Notice the amount of motion in y- and z-direction.

object/cursor closer to me." Others noted caveats and cases where they thought the mirrored interaction would be useful. "In [the] second and third task [ISO tasks] iteration techniques C and D [mirrored] are also easy to use [larger object near the screen]. However A and B feels more natural." (The person who said this did better in the mirrored cases for the static ISO task). "The mirrored tasks were more difficult but when the input wasn't mirrored my interactions occluded the display."

In terms of the skew (perspective; orthographic) differences between the mappings most participants had less to say: "To be honest I couldn't really tell the difference from the previous method" (referring to the two non-mirrored cases). However, some participants—in particular participants with a lot of 3D gaming experience—preferred perspective mapping: "I preferred to have the cursor move in relation to the box" (*i. e.*, perspective; an experienced gamer). This particular comment could mean that the participant preferred the interaction and the visual cues to be consistent.



(a) Front view.



(b) Side view.

Figure 4.12: Trace between two targets in mirrored perspective mapping. Notice the depth confusion.

Others, usually novices, stated a preference for orthographic mapping: "intuitive, felt very much like using a mouse," "felt the most natural."

4.8 Discussion

First, strengths and limitations of the study are discussed to set the context of the statistical results. Following this, each result section is discussed in turn, and then a summary discussion is presented.

4.8.1 Strengths and Limitations

As an exploratory study, the sample consisted of 24 participants. Statistically, when the sample size is small, significance that does exist is less likely to appear as stastically significant in statistical tests. Significance that does show is likely (not guaranteed) to appear more strongly in a large sample size. This last statement depends on the sample being a random sample. In this study, the sample consists mostly of computer science students and alumni. As such, it does not constitute a random sample of computer literate people, nor the larger world. The most probable bias in the sample would be towards the non-mirrored perspective mapping as it is the default mapping and regularly used when something is projected in 3D computer applications and 3D computer games. Yet, this bias is not noticeably evident in the results.

As many depth cues as possible were included, however these were all perspective depth cues. This created an interaction space that is similar to previous 3D interaction research. Thus, the many perspective visual cues support the standard non-mirrored perspective mapping. These visual cues are not only present when interacting in the non-mirrored perspective condition, but also for the three alternative mappings as well. It would be expected that these perspective visual cues would further favour the non-mirrored perspective mapping but this was not seen.

In all three tasks that were statistically analyzed, the participant mean times varied widely both within an individual and between individuals. In terms of individual variance, occasionally participants had difficulty with a selection in what was usually their best mapping. Sometimes, the opposite was true: a person would easily selected a target when overall they had difficulty in that mapping. In terms of variance between individuals, people performed quite differently in the different mappings. That is, some participants consistently performed better in particular mappings in all tasks. Some performed best in both non-mirrored mappings. However, a few performed best in both mirrored mappings. Others performed best in both the orthographic mappings and, likewise, some other participants performed best in both the perspective mappings. A few did well in three of the four mappings. This large variance raises the question as to whether the differences seen would have any practical significance.

4.8.2 Grid Task Completion Time

The Grid task was the first task that participants did and there was no practice time or practice task. This allowed observation of first reaction to the three more unusual mappings. In particular, this is true for the mirrored mapping. Since to our knowledge virtual mirrored spaces have not previously been created, this would be the first experience of mirrored mappings for all participants. Orthographic projections are relatively unusual but are occasionally in use, while perspective mappings are the default for 3D virtual worlds. Thus, for the new alternate mappings, this was the first encounter for the participants. Not surprisingly this first encounter with the mirrored mappings, without visual support, did not make as much sense as non-mirrored mappings for the participants. However, in the experimentation undertaken, the variation is the highest of all the tasks, indicating potentially little or no practical significance. That is, participants varied widely in performance. In fact, some participants consistently performed better in the mirrored mappings than in the non-mirrored mappings.

The differences detected between orthographic and perspective is the first indication that further investigation is needed to determine whether the non-mirrored orthographic mapping, which is in *kinesthetic concordance* (matches our internal body's sense of where objects should be), may be more appropriate for completely novice users and walk-up-and-use scenarios. Another thing to consider is that this is the only task that had the participant interact within the outer edges of the display volume where the perspective distortion has the most effect, and thus the non-distorted orthographic mapping may have the most beneficial impact.

4.8.3 Static ISO Task Completion Time

As the second task a participant had some minimal experience interacting with the system. What is most surprising and interesting is there is no significant difference between one of the mirrored mappings (the perspective mirrored mapping) and both of the non-mirrored mappings. This is a promising indicator that exploration of mirrored spaces is viable. It is uncertain what causes the discrepancy in performance between the mirrored orthographic and mirrored perspective mappings. However, it is possible that this difference lies in the fact that in our daily use of mirrors we are used to relying on visual cues. Although there were no mirrored visual cues, the perspective cues, particularly the strong gridlines, may have guided and assisted the participant in the mirrored perspective mapping.

4.8.4 Moving ISO Task Completion Time

Selection of moving targets is quite different from a static selection [48]. The lack of a significant orthographic versus perspective effect may be because this is the third task and perspective has been learned, or perhaps because moving target acquisition makes the perspective/orthographic changes less relevant. The difference between mirrored and non-mirrored mappings is distinctly less than in the first task. This might suggest moving target selection is a different type of activity, or that participants are learning and improving more in the mirrored mappings than in the non-mirrored mappings and, with more time, participants might perform equally well in the mirrored mappings.

4.8.5 Personal Preference

Participants had clear personal preferences and, surprisingly, most participants preferred the orthographic mapping and not the de facto perspective mapping. In fact, there is a strong positive correlation between novice users and the orthographic mapping preference in our small sample size. As well, the de facto perspective mapping was the least preferred for 3 of the 24 participants. Notice also, that for the number of people (8) that chose perspective mapping as their most preferred mapping first, an equal number chose orthographic mapping as their second most preferred mapping. The mirrored mappings were generally least preferred and this reflects the completion time results. Similarly, as a reflection of poorer performance in the mirrored orthographic mapping in the second task, the mirrored perspective was preferred over the mirrored orthographic.

4.8.6 Trajectory Traces

The example trajectory traces shown above demonstrate examples of particular instances where participants had difficulty in each of the mappings. The mapping in which a particular participant performed best is most often correlated with the case when he/she moved directly to the target. However, even in this best mapping, the participant would have at least one trace that indicated quite a bit of difficulty. As in previous 3D interaction studies [39, 89], participants had the most difficulty in the depth dimension (z). A participant's accuracy was generally high in x and y, however, when a participant was having difficulty selecting a particular target, he/she would be observed to have larger motion in all directions, thus likely decreasing accuracy in x and y in these instances.

4.8.7 Quotes

The quotes, in Section 4.7.7, show the diversity in responses to the various mappings. The mirrored mappings were disliked by some, others enjoyed it (even though they thought it was challenging), and a couple novice users strongly preferred it. Some saw the benefit of using a mirrored display in specific cases or pointed out caveats. The differences between orthographic and perspective were less noticed by participants. Though, there appears to be a split between those with a lot of previous 3D experience (who have a lot of previous experience in the perspective mapping and may be biased due to familiarity) and novice users who felt more at home with the orthographic mapping.

4.8.8 Interaction Style and Strategies

A goal of the study was to observe people understanding and coping with the 3D, and the effect of the mappings, in whatever way made sense to them, or using their own natural process. Each person had a slightly different interaction style and strategy. Some would switch hands to be able to tap targets in the opposite left and right corners. A few with a fair bit of 3D gaming experience made direct movements to roughly where the target was and then made quick jagged movements around the target. Others, in particular novice participants, used a more precise, slower, and deliberate motion. Interestingly, a few would occasionally separate the motion, moving distinctly between the x, y, and z directions. Lastly, one interesting technique in the moving ISO task was to position the cursor above a pillar and wait for the pillar to rise and meet the cursor. Although most participants who tried this technique realized that it is slow and thus did not use it often.

4.8.9 Summary Discussion

The results of the study indicate that all examined mappings are viable and possible alternatives when compared to the de facto standard of perspective mapping. They also show how adaptable humans are to change in control-display mappings and conflict between the visual and sensory-motor models, as previous work suggests [23]. The remarks given by the participants and the ranking results suggest that participants have clear personal mapping preferences and that the standard mapping is not necessarily their preferred first choice. The mean completion times show a noteworthy difference between mirrored and non-mirrored mappings for both the grid and moving ISO task, and for one of the mirrored mappings in the static ISO task. However, for all tasks the variation in mean times is high relative to the actual difference in means, indicating a small practical significance of the effect. The trace analysis indicates that participants had difficulties in all mappings, even in the mappings in which they performed best. Most of these problems occurred in the depth dimension (z) as was to be expected from previous study results [39, 89] but were generally quite accurate in x and y, especially in the ISO tasks. When they had difficulty in z they attempted to adjust, likely causing the observed increased motion and difficulty in height (y) and from time to time in width (x) as well.

Although throughout this study, a perspective box with a grid pattern was used to provide a strong perspective depth cue, in all cases the performance in the orthographic mapping was not meaningfully worse than the perspective mapping. This and the fact that the orthographic mapping was the favourite mapping of the majority of participants is surprising since most 3D rendering and interaction uses perspective projection. It remains to be seen whether a lack of such strong perspective cues or the addition of other depth cues would support the de facto standard perspective mapping or whether orthographic mapping and other skew variations might be more appropriate as the default. The mirrored interactions are new and had no visual support at all in the rendering system (though the perspective component of the mirrored perspective mapping is supported), so they may be even more viable than they appear in this study if appropriate visual support and visual cues were added. Before mirrored visual support can be added research into what would be effective visual support of mirrored mappings would need to be conducted. This open question requires further research.

The differences between results in the different tasks suggest that either certain mappings may be more appropriate for certain tasks or that there is a learning effect. In the first task (grid) participants did best in the non-mirrored orthographic mapping. When reaching the third task (moving ISO) the variation was reduced to a minimum. This could suggest that the non-mirrored orthographic mapping is the easiest to use for the majority of people, and warrants further study. In non-mirrored perspective mapping the grid task was completed slightly slower, in particular for novice users. Even with the visual cues supporting it, participants do approximately as well as when using the non-mirrored orthographic mapping in the moving ISO task. In fact, the visual support for perspective is counterproductive for the orthographic mapping, as the cursor could move behind the perspective walls, and one would think that this would be detrimental to performance in the orthographic mapping. Perhaps the most surprising result is that performance in the mirrored mappings, which have no visual support, improves from the grid task to the moving ISO task and that this improvement is larger than in the non-mirrored mappings. If visual cues to support the mirrored mappings were introduced, it may be that these mappings are even more viable than they appear in this study.

Chapter 5

Conclusion and Future Work

This chapter concludes the thesis by summarizing the research contributions of this work. In the first section, the thesis problems identified in Chapter 1 are reiterated. The next section reiterates the thesis goals, describes the research contributions, and how they address the research goals. Finally, the chapter concludes with suggested future directions.

5.1 Thesis Problems

In Chapter 1, three thesis problems related to 3D Control-Display Mappings were outlined:

1) **What are the components of 3D Control-Display Mappings?** To date, 3D Control-Display Mappings have only been discussed in terms of control-display gain or scale. What are the other components of this relationship?

2) **Can the components of 3D Control-Display Mappings be formalized?** Is there a way to frame and capture the various components into a understandable cohesive whole?

3) **Are any of the alternate 3D Control-Display Mappings viable?** Viable alternative 3D Control-Display Mappings have not been investigated and need to be studied. We need to identify which new possible alternative mappings are

worthy candidates to investigate as viable alternatives. To understand these alternatives and to indicate whether expanding our understanding of Control-Display Mappings will have an impact on 3D Interaction, empirical studies need to be performed.

5.2 Research Contributions

The research presented in this thesis has addressed these problems, and in this section the thesis goals are reiterated and how they were achieved is discussed.

1) **To identify what the components of 3D Control-Display Mappings are:** I identify and describe the components of 3D Control-Display Mappings. *(addressed Problem 1). This goal was addressed in Chapter 3 in Sections 3.1 and 3.2.*

2) **To formalize the discussion of the components of 3D Control-Display Mappings:** I create a conceptual framework and mathematically define the components identified through Goal 1 *(addressed Problem 2). This goal was addressed in Chapter 3*

   To achieve these two goals, in Chapter 3, this work presents the 3D Control-Display Mapping framework. In this framework the components of 3D control-display mappings are delineated. These components are: *scale, order, lag, orientation, alignment, flip, and skew.* Each component is defined mathematically and example mappings are depicted in schematic diagrams. Other than scale, previously, these components have only been minimally explored and thus the 3D control-display framework reveals a large area for potential further exploration and study.

3) **To identify and evaluate potentially viable alternative 3D Control-Display Mappings:** I identify certain issues with the standard mapping and potential mapping alternatives that may address these issues. I perform an initial evaluation of the alternative mappings by designing and running a controlled experiment. I ask participants to perform a series of simple tasks, and to repeat such tasks for four distinct mappings (one traditional, and three alternative). I record and analyze participants' related previous experience, task performance, and mapping preferences. I further discuss and critically analyze the results of the study and the particular choice of the four distinct mappings *(addressed Problem 3) This goal was addressed in Chapter 3 in Section 3.3 and in Chapter 4.*

   In Chapter 3, this work introduces three promising alternative mappings to the de facto standard perspective mapping: orthographic, mirrored perspective, and mirrored orthographic.

   In Chapter 4, this thesis presents the study that I conducted. The four distinct mappings from *Goal 3* were evaluated. All of the mappings appeared viable, and participants did not do meaningfully worse in the orthographic mapping versus the de facto standard perspective mapping and more participants preferred it as well. Participants did surprisingly well in the mirrored mappings, considering they were given no visual cues at all and that this was the first experience of virtual mirrored mappings for all participants, and this opens the way to more investigation into the potential of these mappings and beyond.

5.3 Future Work

The number of studies suggested by the 3D control-display mapping framework is considerable. There is a large area to investigate into how people discover, explore, and understand the 3D space and the mapping between the interactive and virtual space. This study is the first exploratory study. As such, there are many possible future directions. For instance, further investigation into variations and impact of skew mappings could be explored:

1) Investigate whether the orthographic mapping is the most appropriate for novice users and walk-up-and-use scenarios as is hinted at in this study by the participant preference, and performance not being significantly worse than the default mapping.

2) Conduct a similar study on a large display as the effect of skew is more pronounced on a large display. The drift up and towards the center caused by perspective projection appears more drastic on a large screen.

3) Conduct studies with tasks that occur in the edges of the display volume. Linear perspective projection distorts position the most along the edges, and as such the impact of perspective skew is likely the most along the edges.

4) Conduct studies into other variations of skew. In this study, two variations of skew where explored, there are many more variations possible.

Further investigation into flip are possible:

1) Conduct a study on a large display. An initial motivation for exploring mirrored spaces was that people appeared to want to touch and interact directly at the surface of the wall, and tabletop displays in the lab.

2) Conduct studies into other flip variations. Flipping the vertical dimension is sometimes done in mechanical controls, and video games. The effect of this has not been studied.

An alternative avenue of investigation is to explore the addition of visual cues for the novel mappings. As discussed in section 4.8, this study added as many perspective visual cues as possible, as is commonly done in practice to enhance the 3D effect. Here are a few study suggestions:

1) Investigate the addition of visual cues to support mirrored mappings. It is not clear how one would add effective visual cues to support mirrored mappings, and whether they even exist. However effective visual cues have the potential to enhance the viability of the mirrored mappings.

2) Investigate the addition of visual cues to support orthographic mappings, and other skew variations. It is unclear exactly what these visual cues would be.

A full framework has been introduced and the other components in the framework could be studied individually, or in multiple combinations to see how they impact each other. For instance, one could investigate whether scaling techniques, such as the Go-Go technique, are effective in mirrored mappings.

One could study the impact of variations in 3D control-display mappings in other display configurations (particularly the addition of stereopsis), and with other input device types.

This study only looks at the impact of mapping with simple selection and manipulation tasks in the *within reach* metaphor. Further investigation could expand this into more complex selection and manipulation tasks, and beyond into the *at a distance* metaphor. Beyond selection and manipulation, further investigation is warranted into the effect of 3D control-display mappings on navigation tasks.

5.4 Conclusion

In this thesis, the research contributes to the 3D interaction community, by first introducing a 3D control-display mapping framework, and by providing a description of this framework. It includes mappings for scale, order, lag, alignment, orientation, flip, and skew. The effects of skew and flip were explored through the study on four mappings: orthographic, mirrored orthographic, mirrored perspective and the de facto standard perspective. The study results indicate that all three non-standard mappings may be considered as viable alternatives. Surprisingly, the de facto perspective mapping was not the favorite mapping for the majority of participants. Even more surprising is that, with no visual support, performance in the orthographic mapping was not meaningfully worse, and that performance in the mirrored mappings steadily improved even though there was also no visual support for these mappings. In summary, the control-display mapping framework and the results from our study of four control-display space mappings together open the door to a new range of exploration into the space of 3D control-display mapping variations.

Bibliography

[1] Active 3d display. http://www.barco.com/virtualreality/en/stereoscopic/active.asp.

[2] Actuality systems http://www.actuality-systems.com/.

[3] Eternal lands. http://www.eternal-lands.com.

[4] Passive 3d display. http://www.barco.com/virtualreality/en/stereoscopic/passive.asp.

[5] Real-depth 3d http://www.dti3d.com.

[6] Seereal http://www.seereal.com/.

[7] Sensable. http://www.sensable.com.

[8] The Kids in the Hall. Created by The Kids in the Hall Comedy troupe. Prod. Lorne Michaels, CBC, 1988-1994.

[9] Research directions in virtual environments: report of an nsf invitational workshop, march 23-24, 1992, university of north carolina at chapel hill. *SIGGRAPH Comput. Graph.*, 26(3):153–177, 1992. Chairman-Gary Bishop and Chairman-Henry Fuchs.

[10] Maneesh Agrawala, Denis Zorin, and Tamara Munzner. Artistic Multiprojection Rendering. In *Proc. of EGWR 2000*, pages 125–136, London, UK, 2000. Springer-Verlag.

[11] Ravin Balakrishnan. Why Aren't We Using 3D User Interfaces, and Will We Ever? In *Proc. of 3DUI 2006*, page 141, Los Alamitos, CA, March 2006. IEEE

Computer Society.

[12] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia.* Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[13] Renaud Blanch, Yves Guiard, and Michel Beaudouin-Lafon. Semantic Pointing: Improving Target Acquisition with Control-Display Ratio Adaptation. In *Proc. of CHI 2004*, pages 519–526, New York, 2004. ACM Press.

[14] B. G. Blundell and A.J. Schwarz. The classification of volumetric display systems: characteristics and predictability of the image space. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):66–75, Jan-March 2002.

[15] Richard A. Bolt. put-that-there: Voice and gesture at the graphics interface. In *SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, pages 262–270, New York, NY, USA, 1980. ACM Press.

[16] J. Boritz and K. S. Booth. A Study of Interactive 6 DOF Docking in a Computerized Virtual Environment. In *Proc. of VRAIS 1998*, pages 139–146, Los Alamitos, CA, 1998. IEEE Computer Society.

[17] James Boritz. *The Effectiveness of Three-Dimensional Interaction.* PhD thesis, University of British Columbia, Canada, 1999.

[18] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. An Introduction to 3-d User Interface Design. *Presence: Teleoperators & Virtual Environments*, 10(1):96–108, 2001.

[19] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. *3D User Interfaces: Theory and Practice.* Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.

[20] Glenn Bresnahan, Raymond Gasser, Augustinas Abaravichyus, Erik Brisson, and Michael Walterman. Building a Large-Scale High-Resolution Tiled Rear-Projected Passive Stereo Display System Based on Commodity Components. In Andrew J. Woods, Mark T. Bolas, John O. Merritt, and Stephen A. Benton, editors, *Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems X*, volume 5006, pages 19–30, May 2003.

[21] Arthur W. Brody and Chris Hartman. Painting in Space. In Bernice E. Rogowitz and Thrasyvoulos N. Pappas, editors, *Human Vision and Electronic Imaging VI*, volume 4299, pages 602–614. SPIE, 2001.

[22] John Brosz, Faramarz F. Samavati, M. Sheelagh, T. Carpendale, and Mario Costa Sousa. Single Camera Flexible Projection. In *NPAR '07: Proceedings of the 5th international symposium on Non-photorealistic animation and rendering*, pages 33–42, New York, NY, USA, 2007. ACM Press.

[23] Eric Burns, Sharif Razzaque, Abigail T. Panter, Mary C. Whitton, Matthew R. McCallus, and Jr Frederick P. Brooks. The Hand is Slower than the Eye: A Quantitative Exploration of Visual Dominance Over Proprioception. In *Proc of VR 2005*, pages 3–10. IEEE, 2005.

[24] Lewis Carroll. *Through the Looking Glass and What Alice Found There.* MacMillan and Co., Ltd., London, UK, 1899.

[25] Nelson Siu-Hang Chu and Chiew-Lan Tai. Animating chinese landscape paintings and panorama using multi-perspective modeling. *cgi*, 00:0107, 2001.

[26] James C. Chung. A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams. In *SI3D '92: Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 193–196, New York, NY, USA, 1992. ACM Press.

[27] Patrick Coleman and Karan Singh. RYAN: Rendering Your Animation Nonlinearly projected. In *Proc. of NPAR 2004*, pages 129–138, New York, 2004. ACM Press.

[28] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. The CAVE: Audio Visual Experience Automatic Virtual Environment. *Commun. ACM*, 35(6):64–72, 1992.

[29] M. Czerwinski, M. van Dantzich, G. Robertson, and H. Hoffman. he contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3d. In *Interact*, 1999.

[30] Neil A. Dodgson. Autostereoscopic 3d displays. *Computer*, 38(8):31–36, August 2005.

[31] Lucy E. Dunne and Barry Smyth. Psychophysical elements of wearability. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 299–302, New York, NY, USA, 2007. ACM Press.

[32] Niklas Elmqvist and M. Eduard Tudoreanu. Evaluating the effectiveness of occlusion reduction techniques for 3d virtual environments. In *VRST '06: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 9–18, New York, NY, USA, 2006. ACM.

[33] Jim Foley. VR - Past, Present and Future. In *Virtual Reality 2007*, pages xxi–xxi. IEEE, March 2007.

[34] Scott Frees and G. Drew Kessler. Precise and Rapid Iinteraction Through Scaled Manipulation in Immersive Virtual Environments. In *Proc. of VR 2005*, pages 99–106, 2005.

[35] Tinsley A. Galyean and John F. Hughes. Sculpting: an interactive volumetric modeling technique. In *SIGGRAPH '91: Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, pages 267–274, New York, NY, USA, 1991. ACM Press.

[36] Andrew Glassner. Digital cubism. *IEEE Computer Graphics Applications*, 24(3):82–90, May-Jun 2004.

[37] Andrew S. Glassner. Cubism and cameras: Free-form optics for computer graphics. Msr-tr-2000-05, Microsoft, 2000.

[38] H. Paul Grice. Logic and conversation. *Syntax and Semantics, Speech Acts, ed. by Peter Cole and Jerry L. Morgan. New York: Academic Press*, 3:41–58, 1975.

[39] Tovi Grossman and Ravin Balakrishnan. Pointing at Trivariate Targets in 3D Environments. In *Proc. of CHI 2004*, pages 447–454, New York, 2004. ACM

Press.

[40] Tovi Grossman and Ravin Balakrishnan. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 281–290, New York, NY, USA, 2005. ACM Press.

[41] Tovi Grossman and Ravin Balakrishnan. An Evaluation of Depth Perception on Volumetric Displays. In *Proc. of AVI 2006*, pages 193–200, New York, 2006. ACM Press.

[42] Tovi Grossman and Ravin Balakrishnan. The design and evaluation of selection techniques for 3d volumetric displays. In *UIST '06: Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 3–12, New York, NY, USA, 2006. ACM Press.

[43] Tovi Grossman, Daniel Wigdor, and Ravin Balakrishnan. Multi-finger gestural interaction with 3d volumetric displays. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 61–70, New York, NY, USA, 2004. ACM Press.

[44] Yves Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *The Journal of Motor Behaviors*, 19(4):486–517, 1987.

[45] Ken Hinckley, Randy Pausch, and Dennis Proffitt. Attention and visual feedback: the bimanual frame of reference. In *SI3D '97: Proceedings of the 1997*

symposium on Interactive 3D graphics, pages 121–ff., New York, NY, USA, 1997. ACM Press.

[46] Reynald Hoskinson, Caitlin Akai, Brian Fisher, John Dill, and Barry Po. Causes of depth perception errors in stereo displays. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 164–164, New York, NY, USA, 2004. ACM Press.

[47] Isabelle Hyman. *Brunelleschi in Perspective.* Englewood Cliffs, N.J.: Prentice-Hall, 1974.

[48] James L. Knight. Manual Control and Tracking. In Gavriel Salvendy, editor, *Handbook of Human Factors*, pages 182–218. John Wiley and Sons, 1987.

[49] Miho Kobayashi, Michio Oikawa, Takafumi Koike, Kei Utsugi, Masami Yamasaki, and Satoshi Kitagawa. Character interaction system with autostereoscopic display and range sensor. In *Proceedings of the 2nd IEEE Symposium on 3D User Interfaces (3DUI 2007, March 1011, 2007, Charlotte, North Carolina, USA)*, 2007.

[50] D. K. Lemmerman and Joseph Junior LaViola. An exploration of interaction-display offset in surround screen virtual environments. In *3D User Interfaces, 2007. 3DUI '07. IEEE Symposium on*, pages 9–15, 2007.

[51] Jonathan Levene. A Framework for Non-Realistic Projections. Master's thesis, Massachusetts Institute of Technology, 1998.

[52] I. Scott MacKenzie and William Buxton. Extending fitts' law to two-dimensional tasks. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 219–226, New York, NY, USA, 1992. ACM Press.

[53] I. Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. Accuracy Measures for Evaluating Computer Pointing Devices. In *Proc. of CHI 2001*, pages 9–16, New York, 2001. ACM Press.

[54] I. Scott MacKenzie and Stan Riddersma. Effects of Output Display and Control-Display Gain on Human Performance in Interactive Systems. *Behaviour & Information Technology*, 13(5):328–337, 1994.

[55] K. Mason, M. S. T. Carpendale, and B. Wyvill. Projections in context. In *Proceedings of the Western Computer Graphics Symposium*, pages 40–50, 2003.

[56] K. Mason, M.S.T. Carpendale, P. MacMurchy, and B. Wyvill. Eastern perspectives and the question of personal expression. Technical Report 2001-696-19, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada, 2001.

[57] A. D. Milner and M. A. Goodale. *The Visual Brain in Action*, volume 2 of *Oxford Psychology Series.* Oxford University Press, New York, 1995.

[58] Mark R. Mine, Jr. Frederick P. Brooks, and Carlo H. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graph-*

ics and interactive techniques, pages 19–26, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[59] Jurriaan D. Mulder and Robert van Liere. Fast perception-based depth of field rendering. In *VRST '00: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 129–133, New York, NY, USA, 2000. ACM Press.

[60] Jae-Hyeung Park, Yunhee Kim, Joohwan Kim, Sung-Wook Min, and Byoungho Lee. Three-dimensional display scheme based on integral imaging with three-dimensional information processing. *Opt. Express*, 12(24):6020–6032, 2004.

[61] L.M. Parson. Inability to Reason About an Object's Orientation Using an Axis and Angle of Rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 6(6):1259–1277, 1995.

[62] Jeffrey S. Pierce, Brian C. Stearns, and Randy Pausch. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *I3D '99: Proceedings of the 1999 symposium on Interactive 3D graphics*, pages 141–145, New York, NY, USA, 1999. ACM Press.

[63] B. A. Po, B. D. Fisher, and K. S. Booth. Pointing and Visual Feedback for Spatial Interaction in Large-Screen Display environments. In *Proc. of Smart Graphics 2003*, volume 2733 of *Lecture Notes in Computer Science*, pages 22–38, Berlin, Heidelberg, 2003. Springer-Verlag.

[64] E. C. Poulton. *Tracking Skill and Manual Control*. Academic Press, New York, 1974.

[65] I. Poupyrev and E. Kruiff. 20th Century 3DUI Bib: Annotated Bibliography of 3D User Interfaces of the 20th Century. 3rd revision, 2000.

[66] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In *UIST '96: Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 79–80, New York, NY, USA, 1996. ACM Press.

[67] Ivan Poupyrev and Tadao Ichikawa. Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing*, 10:19–35, February 1999.

[68] Ivan Poupyrev, Suzanne Weghorst, and Sidney Fels. Non-Isomorphic 3D Rotational Techniques. In *Proc. of CHI 2000*, pages 540–547, New York, NY, USA, 2000. ACM Press.

[69] G. Rizzolati, L. Fadiga, L. Fogassi, and V. Gallese. The Space Around Us. *Science*, 277:190–191, 1997.

[70] P. Rokita. Generating Depth-of-Field Effects in Virtual Reality Applications. *Computer Graphics and Applications*, 16(2):18–21, March 1996.

[71] Christopher Schmandt. Spatial Input/Display Correspondence in a Stereoscopic Computer Graphic Work Station. In *SIGGRAPH '83: Proceedings of the 10th annual conference on Computer graphics and interactive techniques*, pages 253–261, New York, NY, USA, 1983. ACM Press.

[72] R. Shilling and B. Sinn-Cunningham. Virtual auditory displays, 2002.

[73] Ben Shneiderman. Why Not Make Interfaces Better than 3D Reality? *Computer Graphics and Applications*, 23(6):12–15, November/December 2003.

[74] R. William Soukoreff and I. Scott MacKenzie. Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789, December 2004.

[75] Anthony Steed and Mel Slater. 3d interaction with the desktop bat. In *Computer Graphics Forum*, volume 14, pages 97–104, 1995.

[76] Richard Stoakley, Matthew J. Conway, and Randy Pausch. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proc. of CHI 1995*, pages 265–272, New York, 1995. ACM Press/Addison-Wesley Publishing Co.

[77] Colin Swindells, Barry A. Po, Ima Hajshirmohammadi, Brian Corrie, John C. Dill, Brian D. Fisher, and Kellogg S. Booth. Comparing CAVE, Wall, and Desktop Displays for Navigation and Wayfinding in Complex 3D Models. In *Proc. of CGI 2004*, pages 420–427, Los Alamitos, CA, 2004. IEEE Computer Society.

[78] Melanie Tory, Torsten Moller, M. Stella Atkins, and Arthur E. Kirkpatrick. Combining 2d and 3d views for orientation and relative position tasks. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80, New York, NY, USA, 2004. ACM.

[79] Mario F. Triola. *Elementary Statistics*. Addison Wesley Longman, INC., 1997.

[80] Scott Vallance and Paul Calder. Multi-perspective images for visualisation. In *VIP '01: Proceedings of the Pan-Sydney area workshop on Visual information processing*, pages 69–76, Darlinghurst, Australia, Australia, 2001. Australian Computer Society, Inc.

[81] Lode Vanacken, Tovi Grossman, and Karin Coninx. Exploring the effects of environment density and target visibility on object selection in 3d virtual environments. In *IEEE Symposium on 3D User Interfaces*, pages 117–124, 2007.

[82] Yanqing Wang and Christine L. MacKenzie. Object Manipulation in Virtual Environments: Relative Size Matters. In *Proc. of CHI 1999*, pages 48–55, New York, 1999. ACM Press.

[83] Leonard C. Wanger, James A. Ferwerda, and Donald P. Greenberg. Perceiving Spatial Relationships in Computer-Generated Images. *IEEE Computer Graphics and Applications*, 12(3):44–58, May 1992.

[84] David J. Ward, Alan F. Blackwell, and David J. C. MacKay. Dasher—A data entry interface using continuous gestures and language models. In *Proc. UIST*, pages 129–137. ACM Press, 2000.

[85] Colin Ware. *Information Visualization: Perception for Design*. Academic Press, San Diego, California, 2000.

[86] Colin Ware and Ravin Balakrishnan. Reaching for objects in vr displays: lag and frame rate. *ACM Trans. Comput.-Hum. Interact.*, 1(4):331–356, 1994.

[87] Colin Ware and Glenn Franck. Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions. *ACM Trans. Graph.*, 15(2):121–140, 1996.

[88] Daniel Wigdor, Chia Shen, Clifton Forlines, and Ravin Balakrishnan. Effects of Display Position and Control Space Orientation on User Preference and Performance. In *Proc. of CHI 2006*, pages 309–318, New York, 2006. ACM Press.

[89] S. Zhai, P. Milgram, and A. Rastogi. Anisotropic Human Performance in Six Degree-of-Freedom Tracking: An Evaluation of Three-Dimensional Display and Control Interfaces. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 27(4):518–528, July 1997.

[90] Shuman Zhai. *Human Performance in Six Degree of Freedom Input Control*. PhD thesis, University of Toronto, 1995.

[91] Shumin Zhai. User Performance in Relation to 3D Input Device Design. *SIGGRAPH Computer Graphics*, 32(4):50–54, November 1998.

[92] Shumin Zhai, William Buxton, and Paul Milgram. The “Silk Cursor”: Investigating Transparency for 3D Target Acquisition. In *Proc. of CHI 1994*, pages 459–464, New York, 1994. ACM Press.

Appendix A

Study Materials

In this appendix the following are included: consent form, pre-session questionnaire, and post-session questionnaire. For the post session questionnaire, the first page was given out after the first mapping was complete, the second after the second mapping, and so on, and the last page was completed after all tasks in all mappings were performed.

UNIVERSITY OF CALGARY

Name of Researcher, Faculty, Department, Telephone & Email:

Jeroen Keijser, Master's Student, Department of Computer Science, (403) 210-9499, keijser@cpsc.ucalgary.ca

Supervisor:

Sheelagh Carpendale, Associate Professor, Department of Computer Science, (403) 220-6055, sheelagh@cpsc.ucalgary.ca

Title of Project:

3D Interaction Techniques on Large Displays

This consent form, a copy of which has been given to you, is only part of the process of informed consent. If you want more details about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this research study.

Purpose of the Study:

We are currently investigating how 3D interaction with large displays is affected by the mapping between the real 3D physical world and the virtual world represented on a large display. We are investigating the effect these mappings have on 3D interactions in order to better understand how to support 3D interaction on large displays.

What Will I Be Asked To Do?

You will be asked to interact with objects in a 3D virtual space on large displays. Your tasks may involve moving digital objects to different locations, navigating through a virtual space, and making a simple computer model of an everyday object.

We will be observing and programmatically capturing your actions, as well as videotaping you during the course of the session. This videotaping is optional and you can still participate if you choose not to be videotaped. You will also be asked to complete a pre- and post-session questionnaire to further our investigation. It is estimated that your involvement will take approximately one hour, and you will be remunerated for your time.

Participation in this study is completely voluntary and you may refuse to participate altogether or at any time during the study without any repercussions.

What Type of Personal Information Will Be Collected?

Should you agree to participate, you will be asked to provide both your gender and age. No other personal identifying information will be collected in this study, and all participants shall remain anonymous.

There are several options for you to consider if you decide to take part in this research. You can choose all, some or none of them. Please put a check mark on the corresponding line(s) to grant me your permission to:

I grant permission to be audio taped:	Yes: ___ No: ___
I grant permission to be videotaped:	Yes: ___ No: ___
I grant permission to have video or still images of me used in publications and/or presentations:	Yes: ___ No: ___

Are there Risks or Benefits if I Participate?

There are no known risks involved in participating in this experiment. You will be given $10 as compensation for your time. If you decide for any reason to withdraw from the experiment, you will still receive this compensation.

What Happens to the Information I Provide?

Participation is completely voluntary and confidential. Note that because you may be working with others to complete the experiment, the participants in the group besides yourself may be able to identify you. Besides this limitation, you will remain completely anonymous. You are free to discontinue participation at any time during the study. If you decide to discontinue participation, data collected up to that point will be destroyed. No one except the researcher and his supervisor will be allowed to see or hear any of the answers to the questionnaire or the interview tape. There are no names on the questionnaire. Only group information will be summarized for any presentation or publication of results. The questionnaires are kept in a locked cabinet only accessible by the researcher and his supervisor. The anonymous data will be stored for five years on a computer disk, at which time, it will be permanently erased.

Signatures (written consent)

Your signature on this form indicates that you 1) understand to your satisfaction the information provided to you about your participation in this research project, and 2) agree to participate as a research subject.

In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from this research project at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: (please print) ____

Participant's Signature ____Date: _______________

Researcher's Name: (please print) ____

Researcher's Signature: ____Date: ________________

Questions/Concerns

If you have any further questions or want clarification regarding this research and/or your participation, please contact:

Mr. Jeroen Keijser and Dr. Sheelagh Carpendale
Department of Computer Science
403-210-9499, {keijser,sheelagh}@cpsc.ucalgary.ca

If you have any concerns about the way you've been treated as a participant, please contact Patricia Evans, Associate Director, Research Services Office, University of Calgary at (403) 220-3782; email plevans@ucalgary.ca

A copy of this consent form has been given to you to keep for your records and reference. The investigator has kept a copy of the consent form.

UNIVERSITY OF CALGARY

Pre-Session Questionnaire

3D Interaction Techniques on Large Displays

A. Circle the number that best indicates how long you have used computers for:

1	**2**	**3**	**4**	**5**
no experience	less than a month	less than a year	1 to 5 years	more than 5 years

Comments: ______________________________

____

B. Circle the number that best indicates your weekly use of computers:

1	**2**	**3**	**4**	**5**
none	0-5 hours/week	5-15 hours/week	15-35 hours/week	> 35 hours/week

Comments: ______________________________

____

C. Circle the number that best indicates your experience with large displays (tabletop displays, dome displays, over 30 inch computer displays):

1	**2**	**3**	**4**	**5**
no experience	used once	used 2-5 times	used 5-20 times	used >20 times

Comments: ______________________________

____

D. Circle the number that best indicates your experience with 3D computer graphics / 3D gaming :

1	**2**	**3**	**4**	**5**
no experience ----------------	play rarely a few times a year	play moderately few times a month	play regularly more than once a week	keen, play a lot play most days

Comments: ______________________________

____

E. Circle the number that best indicates your experience with 3D in general (3D art, perspective drawing, sculpture, higher dimensional math/ physics etc.) :

1	**2**	**3**	**4**	**5**
general knowledge	occasional 3D thinking	some 3D thinking	think comfortably in 3D	work in 3D

Comments: ____

____

F. What is your gender (for statistical purposes)?

[ ] Male **[ ] Female**

G. Are you primarily right- or left-handed?

[ ] Right-handed **[ ] Left-handed**

UNIVERSITY OF CALGARY

Post-Session Questionnaire

3D Interaction Techniques on Large Displays

Interaction technique A

A. Circle the number that best indicates how easy to **use** you found the Interaction Technique just experienced:

1	2	3	4	5
very difficult				very easy

Comments: ______________________________

____

B. Circle the number that best indicates how easy to **learn** you found the Interaction Technique just experienced:

1	2	3	4	5
very difficult				very easy

Comments: ______________________________

____

C. Circle the number that best indicates how much you **enjoyed** using this Interaction Technique:

1	2	3	4	5
not at all				very much

Comments: ______________________________

____

D. Circle the number that best indicates your impression of how **well** you completed the tasks using this Interaction Technique:

1	2	3	4	5
not well at all				very well

Comments: ______________________________

____

UNIVERSITY OF CALGARY

Post-Session Questionnaire

3D Interaction Techniques on Large Displays

Interaction technique B

A. Circle the number that best indicates how easy to **use** you found the Interaction Technique just experienced:

1	2	3	4	5
very difficult				very easy

Comments: ______________________________

B. Circle the number that best indicates how easy to **learn** you found the Interaction Technique just experienced:

1	2	3	4	5
very difficult				very easy

Comments: ______________________________

C. Circle the number that best indicates how much you **enjoyed** using this Interaction Technique:

1	2	3	4	5
not at all				very much

Comments: ______________________________

D. Circle the number that best indicates your impression of how **well** you completed the tasks using this Interaction Technique:

1	2	3	4	5
not well at all				very well

Comments: ______________________________

UNIVERSITY OF CALGARY

Post-Session Questionnaire

3D Interaction Techniques on Large Displays

Interaction technique C

A. Circle the number that best indicates how easy to **use** you found the Interaction Technique just experienced:

1	**2**	**3**	**4**	**5**
very difficult				very easy

Comments: ____

____

B. Circle the number that best indicates how easy to **learn** you found the Interaction Technique just experienced:

1	**2**	**3**	**4**	**5**
very difficult				very easy

Comments: ____

____

C. Circle the number that best indicates how much you **enjoyed** using this Interaction Technique:

1	**2**	**3**	**4**	**5**
not at all				very much

Comments: ____

____

D. Circle the number that best indicates your impression of how **well** you completed the tasks using this Interaction Technique:

1	**2**	**3**	**4**	**5**
not well at all				very well

Comments: ____

____

UNIVERSITY OF CALGARY

Post-Session Questionnaire

3D Interaction Techniques on Large Displays

Interaction technique D

A. Circle the number that best indicates how easy to **use** you found the Interaction Technique just experienced:

1	2	3	4	5
very difficult				very easy

Comments: ____

____

B. Circle the number that best indicates how easy to **learn** you found the Interaction Technique just experienced:

1	2	3	4	5
very difficult				very easy

Comments: ____

____

C. Circle the number that best indicates how much you **enjoyed** using this Interaction Technique:

1	2	3	4	5
not at all				very much

Comments: ____

____

D. Circle the number that best indicates your impression of how **well** you completed the tasks using this Interaction Technique:

1	2	3	4	5
not well at all				very well

Comments: ____

____

UNIVERSITY OF CALGARY

Post-Session Questionnaire

Ranking:

I. Based on your experience with the various Interaction techniques, please rank the following in order of your preference for the tasks performed, with 1 as your **most** preferred, and 4 as your **least** preferred:

[ ] **Interaction technique A**
[ ] **Interaction technique B**
[ ] **Interaction technique C**
[ ] **Interaction technique D**

II. Do you have any comments regarding your preferences?

III. Do you have any other comments regarding the study?

Appendix B

Ethics Approval

UNIVERSITY OF
CALGARY

CERTIFICATION OF INSTITUTIONAL ETHICS REVIEW

This is to certify that the Conjoint Faculties Research Ethics Board at the University of Calgary has examined the following research proposal and found the proposed research involving human subjects to be in accordance with University of Calgary Guidelines and the Tri-Council Policy Statement on *"Ethical Conduct in Research Using Human Subjects"*. This form and accompanying letter constitute the Certification of Institutional Ethics Review.

File no:	**4448**
Applicant(s):	**Jeroen Keijser** M. Sheelagh Carpendale Tobias Isenberg Mark Hancock
Department:	**Computer Science**
Project Title:	**3D Interaction Techniques on Large Displays**
Sponsor (if applicable):	

Restrictions:

This Certification is subject to the following conditions:

1. Approval is granted only for the project and purposes described in the application.
2. Any modifications to the authorized protocol must be submitted to the Chair, Conjoint Faculties Research Ethics Board for approval.
3. A progress report must be submitted 12 months from the date of this Certification, and should provide the expected completion date for the project.
4. Written notification must be sent to the Board when the project is complete or terminated.

Janice Dickin, Ph.D, LLB,
Chair
Conjoint Faculties Research Ethics Board

Originally issued: 20 June 2005
Revised: 14 March 2006

Distribution: (1) Applicant, (2) Supervisor (if applicable), (3) Chair, Department/Faculty Research Ethics Committee, (4) Sponsor, (5) Conjoint Faculties Research Ethics Board (6) Research Services.

2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4 • www.ucalgary.ca

Appendix C

Co-Author Permission

UNIVERSITY OF CALGARY

July 17, 2007

University of Calgary
2500 University Drive NW
Calgary, Alberta
T2N 1N4

Permission for the Use of

Jeroen Keijser, Sheelagh Carpendale, Mark Hancock, and Tobias Isenberg. Exploring 3D Interaction in Alternate Control-Display Space Mappings. *In Proceedings of the 2nd IEEE Symposium on 3D User Interfaces* (3DUI 2007, March 10–11, 2007, Charlotte, North Carolina, USA). Los Alamitos, CA. IEEE Computer Society, pages 17-24, 2007.

I, Sheelagh Carpendale, give Jeroen Keijser permission to use co-authored work from our paper listed above in his thesis dissertation and to have this work microfilmed.

Sheelagh Carpendale — July 17, 2007 — Date

I, Mark Hancock, give Jeroen Keijser permission to use co-authored work from our paper listed above in his thesis dissertation and to have this work microfilmed.

Mark Hancock — July 18, 2007 — Date

I, Tobias Isenberg, give Jeroen Keijser permission to use co-authored work from our paper listed above in his thesis dissertation and to have this work microfilmed.

Tobias Isenberg — July 17, 2007 — Date

2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4 • www.ucalgary.ca