Testing the Two-Stream Hypothesis in an Immersive Virtual Environment

Rajkumar Parasuraman Viswanathan

Department of Electrical & Computer Engineering
McGill University
Montréal, Canada

April 2013

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Masters in Engineering.

2013/04/15

Abstract

A great deal of behavioural research has gone into a proposed distinction between two separate streams for visual processing, vision for action and vision for perception. Research on perceptual and geometric illusions has gone a long way in determining this proposed dissociation in visual processing. These illusions fool the brain into misjudging object sizes but at the same time do not affect fingers from scaling to the correct size while grabbing. This effect is maintained even when the stimuli are three-dimensional. The mechanisms mediating the visual control of object-oriented actions are thought to operate in egocentric coordinates. We would therefore like to know whether this effect is maintained when reaching for the illusion with a virtual arm, where there is an indirect pairing of visual and proprioceptive feedback, a process essential for pairing the external visual scene onto egocentric coordinates. Our research shows that while the two stream effect is maintained in a real world, it is lost out in a virtual world where there is a lack of haptic feedback. It is also seen that participants underestimate depth unless given an external feedback, in our case, a change of colour in the virtual arm within a virtual environment with a depth grid.

Sommaire

Une grande partie de la recherche comportementale consiste d'une distinction proposée entre deux flux distincts pour le traitement visuel, la vision pour l'action et la vision pour la perception. La recherche sur les illusions perceptives et géométriques a parcouru un long chemin vers la détermination de cette proposition de la dissociation du traitement visuel. Ces illusions trompent le cerveau à mal juger les tailles des objets, tout sans empêchant les doigts de mettre à l'échelle les tailles correctes alors qu'ils saisissent ces objects. Cet effet persiste même quand les stimuli sont en trois dimensions. Les mécanismes qui facilitent le contrôle visuel des actions sur objects sont soupçonnés de fonctionner en coordonnées égocentriques. Nous aimerions donc savoir si cet effet persiste quand on essait d'atteindre l'illusion en utilisant bras virtuel, où si il existe un couplage indirect entre le visuel et la rétroaction proprioceptive, un processus essentiel pour la superimposition de la scène visuelle externe sur les coordonnées égocentriques. Notre recherche montre que, bien que l'effet de deux flux est maintenu dans le monde réel, il est perdu dans un monde virtuel où la rétroaction haptique est absente. Il est également observé que les participants sous-estiment la profondeur, à moins qu'une rétroaction externe est donnée, comme, dans notre cas, un changement de couleur dans le bras virtuel dans un environnement virtuel avec une grille de profondeurs.

Acknowledgments

First of all, I would like to sincerely extend my gratitude and appreciation to my supervisor, Dr. Jeremy Cooperstock, for his guidance and discussions over the course of my masters program. His valuable suggestions helped me conduct my experiments in a formal way and also write my masters thesis. I would like to thank SRE lab members GuangYu Wang, Dalia El-Shimy and Stephane Pelletier for helping me with camera calibration and the motion capture system. I would also like to thank Dr. Wissam Musallam and Dr. Amir Schmuel whose course in Neural Prosthetics helped me gain insight into the functioning of the brain which was a huge part of this research. Special thanks goes to Dr. Ian Gold and Alireza Hashemi for entrusting this research to me. A word of thanks for Alireza for his help in data analysis and the neuropsychological point of view of the experimental results. The experiments would not have been possible without active participation from SRE lab members, CIM members and other students from McGill University. Thanks to all those who actively participated in the experiments. Finally, I would like to dedicate this thesis to my parents - Mrs. Usha Viswanathan and Mr. Viswanathan - my brother - Mr. Sanjay Kumar - and to all my family and friends in Montreal and Chennai for all the support and help they have given me throughout my life. A special thanks to Ms. Mirunalini Thirugnanasambandam for her help and support during my masters program. Funding for this project was provided by the Natural Sciences and Engineering Research Council (NSERC) and the Networks of Centres of Excellence on Graphics, Animation and New Media (GRAND).

Contents

List of Figures

List of Tables

List of Acronyms

FOV	Field Of View
HMD	Head-mounted Display
IPD	Interpupillary Distance
IPT	Immersive Projection Technology
IVE	Immersive Virtual Environment
OSC	Open Sound Control
RFI	Rod and frame illusion
SID	Spatially Immersive Display
STI	Simultaneous tilt illusion
VR	Virtual Reality

Chapter 1

Introduction

Everyday human activities involve heavy use of perception of objects through vision and corresponding action based on vision. Perception can be defined as the brain's ability to organize, identify and interpret sensory information to produce a mental image of an object. The human brain processes vision in the visual cortex, also called the striate cortex or V1 region of the brain located in the occipital lobe, in the back of the brain. This region, which is the primary visual cortex, is divided into two parts, one in the left and and the other in the right hemisphere of the brain. The two cortices receive visual signals from the opposite visual fields, i.e., visual signals flow from the left visual field to the visual cortex in the right hemisphere and vice-versa. Thus, there are two V1s located within the brain, each on one hemisphere. It is believed that each V1 transmits information through two corticulo-corticular pathways called the dorsal and ventral streams. The **dorsal stream**, also called the "how" or "where" pathway processes object-oriented action while the **ventral stream**, called the "what" pathway processes perception and object attributes. The idea that two pathways process visual information was first defined by Ungerleider and Mishkin [1] and is called the "two-stream hypothesis". However, this idea has been heavily contested by the one-stream hypothesis where only one pathway processes visual information i.e., both perception and action. It is believed from the one-stream hypothesis that action follows as a result of perception.

Perception involves identifying the object and its various attributes such as shape, or colour with the requirements that the person is able to see the object and has previous knowledge of its attributes to identify them. To perform action on an object, it is necessary that the brain calculates the distance of that object away from the human body. The human vision system uses depth perception to locate identified objects in space. Depth perception or stereoscopic vision is the ability to identify the distance between the point of view and an object in the field of view. The human visual system provides depth perception by processing views from both eyes and triangulating the distance to an object being viewed. If the distance to an object is d and the interocular distance [1] is l, then Equation 1.1 shows that d can be calculated as

$$d = \frac{l \times \sin(\alpha) \times \sin(\beta)}{\sin(\alpha + \beta)} \tag{1.1}$$

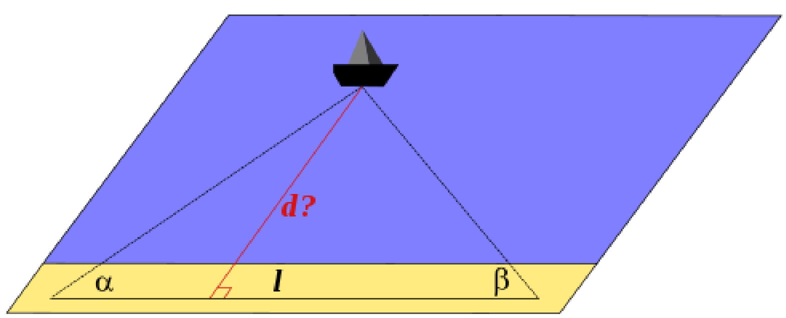

Figure 1.1 Triangulation method which is used to compute the distance of an object

1. Distance between the centre of the eyes

Figure 1.1 shows the triangulation method where d is the depth of the object, and l the distance between two points. Thus, depth is inferred using binocular vision where α and β are the angles between each eye and the object. Depth can also be inferred by monocular vision with the help of many cues such as motion parallax, perspective, relative size, depth from motion and occlusion. This is predominant in most animals where each eye is located on either side of the head and do not view the same object simultaneously. In humans and primates, the eyes are located in front of the head and hence binocular vision is predominantly used for depth perception. In addition to this, human beings also move their eyes so the optical axes converge at the point where the object is located as shown in Figure 1.1.

Gestalt laws of organization are often applied to visual perception. There are six main factors that determine how the visual system perceives things namely, proximity, closure, similarity, symmetry, common fate and continuity. Some of these factors, primarily closure, are used for creating illusions.

1.1 Illusions

Illusions have long been part of scientific studies in the fields of neuropsychology, computer vision and also in arts. Illusions are generally of two types, either physiological or cognitive. Physiological illusions appear due to sudden or increased competing stimuli of a specific type to the eyes. The theory behind physiological illusions is that a stimulus inhibits or causes physiological imbalance when it is repeatedly exposed. A quite famous illusion is the Hermann grid illusion shown in Figure 1.2. The grid consists of black squares intersected by thin white lines. The attributes of the grid such as shape and colour create the illusion that there are black dots present in the intersection spaces of perpendicular white lines.

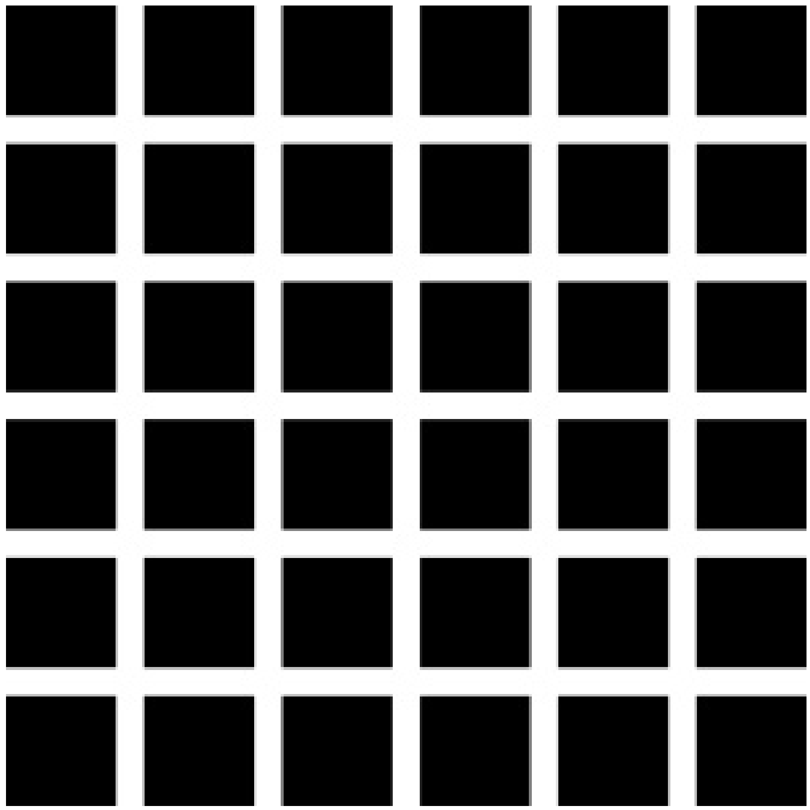

Figure 1.2 The Hermann Grid - The intersection of perpendicular white lines with contrasting black spaces produce an illusion of black dots appearing

The second type of illusions, called cognitive illusions, occur due to assumptions made by the human brain leading to "unconscious inferences" as concluded by Hermann Von Helmholtz. These are the kind of illusions that are primarily used in arts. A prime example would be illusions where one sees a few human faces within a scenery that contains a landscape. Cognitive illusions are further subdivided into four kinds, namely ambiguous illusions, distorting illusions, paradox illusions and fictions. Ambiguous illusions are those where the brain switches between two outlines. In other terms, a single image is perceived in different ways. The Rubin vase is a famous example of this kind where the human brain perceives the outline of a vase and also the lateral view of two human faces facing each other. Another illusion of this kind is an image of one half of a human face. At times, the face seems to be looking at the side while at other times, it appears to be facing front on. These kinds of illusions are generally created by the negative space surrounding a figure. Paradox illusions are created by objects that are impossible to construct such as the impossible staircase and the Penrose triangle. For instance, in the staircase model, the illusion is created by the staircase making four 90-degree turns and thus creating an infinite loop of ascent and descent. Fictions are illusions that are created when the brain perceives presence of objects even though they are not in the stimulus. Distorting illusions are created by using the geometry of objects such as size, shape, length, position or curvature. Of these four, distorting illusions, also referred to as geometrical-optical illusions, are used for studying the two-stream hypothesis. The reasoning behind this is that these visual illusions tend to affect perception and fool the brain, particularly when there is a lack of prior knowledge about the illusion. Additionally, actions such as grabbing the illusion can be performed on these. The question is whether this performed action, such as grabbing the illusion, is also fooled by these illusions. Grabbing the illusion in general refers to grasping the boundaries of an object within the illusion where the object is one whose size is perceived to be different because of its surroundings. While some studies argue that action is not affected by illusions, there are others that argue against it. Tests on users' perception and action on these illusions and their subsequent results could possibly add to or reduce weight from the idea of a two-stream hypothesis. Some commonly used illusions to study this phenomenon are the Judd illusion and Muller-Lyon illusion seen in Figure 1.3, the Ebbinghaus illusion, the simultaneous-tilt illusion, the rod and frame illusion and the induced displacement effect.

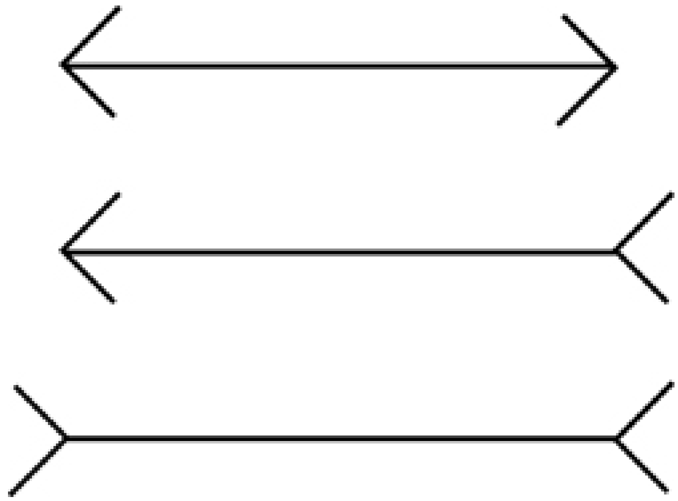

Figure 1.3 Muller Lyon illusion - The lines are of the same size but appear to be different in sizes due to orientation of the arrows

The Ebbinghaus illusion is probably the most popular among these illusions to study the two-stream hypothesis. Lots of studies have been conducted on the Ebbinghaus illusion including the two-stream hypothesis, on how the illusion is created and the factors that affect the illusion. For the purpose of our experiments, we use the Ebbinghaus illusion which is described in detail below.

The Ebbinghaus illusion consists of two central spheres of the same radius each surrounded by a set of spheres. One of the spheres is surrounded by smaller spheres while the other sphere is surrounded by bigger spheres. This creates an optical illusion where the central sphere on the left, seen in Figure 1.4, seems to be bigger than the central sphere on the right. The illusion is generally believed to be created by the presence of the surrounding spheres and their respective sizes. Studies by Haffendan et al. [2] show that the illusion is created by the distance between the central sphere and the surrounding spheres in addition to the difference in their sizes.

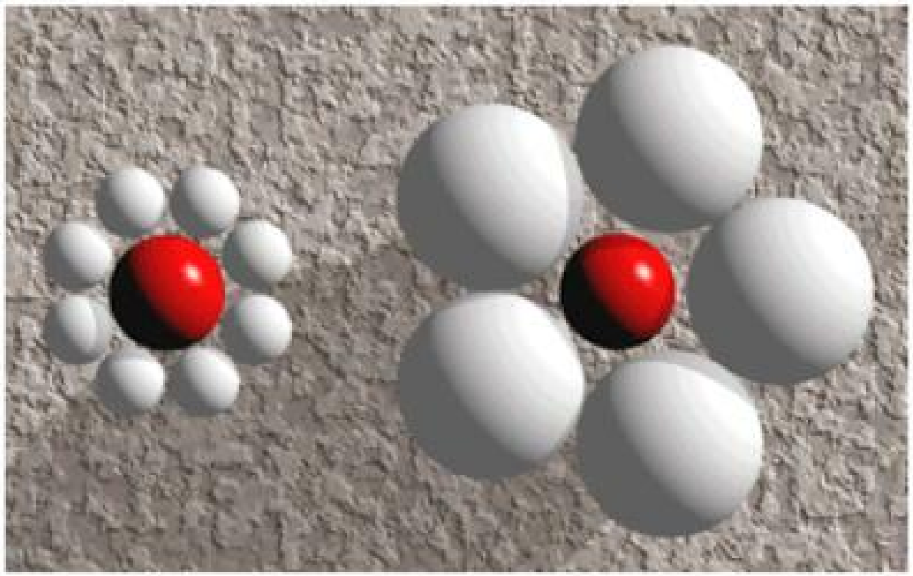

Figure 1.4 Ebbinghaus Illusion - The centre circles appear to be different in size although they are of the same dimensions

Massaro and Anderson [3] did a study of the Ebbinghaus illusion where they determine the comparative nature of the illusion. Centre circles are surrounded by context circles, which provide the standards based on which the size of the center circle is judged. There are various factors that affect the Ebbinghaus illusion such as the sizes of the context circles and the difference in sizes between the smaller and larger context circles, the distance between the context circles and the centre circles, the number of context circles surrounding the centre circle and also lighting contrast between context circles and the centre circle. Research [2] has shown that the illusion grows with distance between the context circles and the centre circles. In their experiments, the smaller context circles and the centre circle had two distance separations. One distance separation had a finger width separation between centre circle and the smaller context circles. This distance was the same as the other centre circle to the larger context circles. They termed this arrangement as adjusted small. The other distance separation had close to no separation between the centre circle and the smaller context circle, called traditional small. They found that when the distance between the context circles and the centre circles was same for both smaller and larger context circles, grasp scaling difference [2] was very low around 0.21 mm compared to the manual estimation [3] results which was around 2.65 mm. In the case of the traditional small and traditional large context circles, grasp scaling difference was around 1.2 mm while manual estimation differences were close to 3.5 mm. They also found that every 1 mm increment in target diameter resulted in a 1.85 mm increment in manual estimation while grasp scaling was affected only by 0.88 mm. In conclusion, they found that grasp scaling differences were a function of the distance between the centre circle and the inner edge of the context circles. Research also shows that the effect of the illusion increases with an increase in the number of context circles surrounding the centre circles. An increase in the difference in sizes between the larger and smaller context circles also results in an increase in the effect of the illusion. An increase in the distance between the centre circle and the larger context circles results in an increased underestimation of the radius of the centre circle. Further research also shows that increasing the lightness contrast of either the larger or smaller context circles relative to the centre circle causes the centre circle to appear larger than it is. Jaeger and Grasso [4] studied the effects of contrast and contour in the Ebbinghaus illusion and the relation between the effects of lightness of the contours and the size and location of the context circles in the Ebbinghaus illusion. Their studies showed that the greater the lightness contrast between context circles and the centre circle, the larger the centre circle appears. This can be explained by the fact that the context contours of greater lightness contrast are registered more vigorously by the visual system and hence the centre circle seems to be attracted more to the context circles making it appear larger than it is.

2. Grasp scaling difference refers to difference in grip apertures while grabbing the centre circles.

3. Difference in grip aperture during perception. Subjects were asked to open their index finger and thumb till they felt they had matched the size of the centre circle in question.

Previous work on visual illusions such as the Ebbinghaus illusion and grasping such illusions to study the two-stream hypothesis have mostly been performed in real environments [5, 6, 2, 3]. Our goal is to find out if the effects of the illusion while perceiving and acting on it are maintained within a virtual environment. As mentioned earlier, in a virtual environment, there is an indirect relation between visual and proprioceptive feedback, a factor essential for pairing an external visual scene onto egocentric coordinates in which object-oriented actions operate. Egocentric coordinates have their coordinate system origin in the body and the relative directions of an object in space are obtained with respect to the body. To test the two-stream hypothesis and the differences in perception and action within a virtual world when compared to a real world, we need an Immersive Virtual Environment (IVE).

1.2 Virtual Reality and Immersive Displays

To test the two-stream hypothesis in a virtual world, we require an immersive display where users will see the illusion in the virtual world and will also be able to grab the illusion in the virtual world without being able to see the actual position of their hands. Here, we discuss what factors affect perception and action within an IVE.

Bolas [7] gave a review of the human factors that go into the design of an immersive display. The factors include general usability issues, display technology, optical human factors, data and video interface, navigation and manipulation and tracking. General usability issues include ease of use, support for multiple users and also multiple users working on different tasks simultaneously. Optical human factors include head sizes, interpupillary distances and other individual vision related problems such as myopia or astigmatism.

Steed and Parker [8] discuss 3D selection strategies of Spatially Immersive Displays(SIDs). In an IVE, selecting an object is either achieved through collision between the desired object in VR and the user's virtual hand or through a ray projecting in the direction of the hand and its intersection with the object. The first method is called as virtual hand technique while the the second is referred to as ray casting. Following their previous work in 2004, Steed and Parker [9] also evaluated the effectiveness of interaction techniques in both Immersive Projection Technology (IPT) and HMDs. In particular, they looked into the virtual hand and ray casting techniques. The results of these evaluations helped them give guidelines for selecting the interaction technique used in a IVE. In general, they found that performance was better in IPTs compared to HMDs. However, they also note that a three walled IPT does not produce a completely immersive experience. Thompson et al. [10] talk about the effect of the quality of graphics when judging distances in IVEs. It is well known that distances are usually underestimated in virtual environments compared to real environments. From their experiments, they found that graphics does not affect distance judgements. They also propose that a full sense of presence might help judge virtual distances better. Plumert et al. [11] studied distance perception in real and virtual environments. Their experiments involved time-to-walk estimates over certain distance in both real and virtual environments. Their experiments showed that underestimation happened in both real and virtual worlds with time-to-walk estimates and in addition to that, distance perception could be better in virtual environments involving larger displays (SIDs and IPTs) compared to HMDs. Draper et al. [12] talk about the effects of a head-coupled control and a HMD on large search area tasks. Their findings suggest that HMDs do not offer a huge advantage over traditional displays in large search area tasks. Ruddle et al. [13] tested the differences between desktop displays and HMDs in navigating large scale VR environments. They found that on an average, participants navigated the VR environment twelve percent quicker using the HMD. Head direction changes were also higher in participants when using desktop displays as against HMDs where it was nine percent lesser. It was also found that participants developed a significantly more accurate sense of relative straight-line distance when using a HMD.

Head-Mounted Displays, as the name suggests, are display devices worn on the head and generally are of two types, monocular and binocular depending on whether one or two displays are present in the device. The display screen is mostly made of LCDs, OLEDS or LCOs. The first HMD, called the Sword of Damocles, was created by Ivan Sutherland and Robert Sproull [14]. Two tubes, separated at an interpupillary distance(IPD), which is the distance between the centres of the two eyes, end in CRT displays to deliver images to the two retinas , thus creating a stereoscopic image. This allows users to see a 3D object in a VR scene. This HMD allowed for the user to move three feet off axis in any direction to view objects better and also for a vertical tilt of up to 40 degrees and a horizontal tilt of 360 degrees. One of the major problems in creating truly realistic 3D objects is the "hidden line problem", which is to compute which portions of an object are hidden by another in a VR scene. Considering the technology during this HMD's development period, they used only transparent "wire frame" line drawings. This is considered the predecessor of current binocular HMDs which are widely used in various fields such as gaming, medicine, sports training and aviation.

We had earlier discussed about the idea behind depth perception and how the human vision computes the depth of an object in space. In order to create realistic scenes within a HMD, we need to have a model that allows for users to see stereoscopic images. In a virtual environment within a HMD, orthostereoscopy is defined as the constancy of the perceived size, shape and relative positions as the head moves around [15]. To achieve this, Robinett and Rolland define a computational model for the geometry of a head-mounted display. To calculate this computational model, they take into account among various factors, IPD, screen resolution, position of screen edges, horizontal and vertical fields of view (FOV). Willemsen et al. [16] studied the effects of field of view and binocular viewing restrictions in real world distance perception by creating an environment analogous to one seen in a HMD. Their results show that FOV and binocular viewing restrictions do not cause underestimation of distances generally seen with HMDs. They also consider the possibility of graphics affecting distance perception in virtual environments. Kawara et al. [17] studied object handling in virtual environments within a head-mounted display. Their studies show that subjects required considerably less time when some sort of feedback was given upon completion of a task. In this case, subjects were asked to move three 5 cm diameter circles from left to right and back. One half of the subjects were given an acoustic feedback while the other half did not receive any feedback. On an average, subjects that were given an acoustic feedback took 20 s to complete the task while the rest took 40 s. Thus, they concluded that some sort of sensory feedback is necessary to make HMD systems more "human-friendly" and more usable.

To test the two-stream hypothesis in a virtual world, we had to decide on an effective mode of display that would allow the user to see an immersive virtual environment. Given that we wanted to test the differences in grabbing in the real and virtual worlds, we had to ensure that users were not able to see their physical hand during grabbing in the virtual world. We had discussed the differences between SIDs and HMDs from which the general conclusions were that SIDs were better used in navigating large search areas. It was also suggested that HMDs enable faster navigation compared to SIDs. Since our experiments did not require too much navigation and also were conducted in a small virtual area, we decided on using HMDs for the IVE.

1.3 Literature Review

Ungerleider and Mishkin [1] suggested a two-stream hypothesis for visual processing of objects. Of the two streams, the ventral stream processes object attributes such as shape, size and colour. The location of the object in space and guiding any action on the object such as grabbing or flicking are processed by the dorsal stream. Goodale and Milner [18] published a review on this two-branch visual system hypothesis. According to them, both the cognitive and sensorimotor branches start together from the primary visual cortex. The cognitive branch then goes into the temporal lobe through the ventral stream and the sensorimotor branch goes into the parietal cortex through the dorsal stream.

The initial ideas behind the two-stream hypothesis stemmed from studies conducted on the striate cortex of a monkey from which was proposed the presence of two multisynaptic corticocorticular pathways [1]. In 1969, Schneider [37] proposed an anatomical separation between visual processing of a stimulus and the identification of the stimulus. He attributed the location of the stimulus, or location, to the retinotectal pathway and the identication of the stimulus to the geniculostriate system. Although his original proposal was rejected later with the distinction attributed to the dorsal and ventral pathways, the notion of distinction between visual processing and identification of stimulus remained. Thus, the distinction was made between object identification, “what”, and spatial location of the object, “where”. The distinction in this anatomical separation thus depended on the input distinctions of object attributes and object location. Goodale and Milner’s review claimed that both these streams are simultaneously activated during vision-based action. They proposed that the functional dichotomy between the ventral and dorsal streams was better explained by the “what” versus “how” distinction rather than the “what” versus “where” distinction as previously thought of. In other words, the functional differences between perception and skilled visuomotor action were not observed between object vision and spatial vision.

By way of example, it was found that one particular patient with lesions in the inferotemporal region could not identify objects, their sizes or their orientation. However, the patient was able to grasp the object perfectly when asked to manipulate it. In this case, even though the “what” pathway was non-functional, the patient was able to use the “how” pathway to perform a task. In essence, if there is a dissociation between the “what” and “where” pathway, the functional differences are between object attributes and object location. However, when the dissociation is between the “what” and “how” pathways, the functional differences are between object attributes and how the motor system plans to act on the object. In addition, Goodale and Milner conclude that spatial attention or the originally proposed “where” pathway is physiologically non-unitary and can be associated with both the ventral and dorsal streams.

Experiments conducted on patients with lesions in certain regions of the brain further enhanced the notion of two-stream visual processing. Patients with lesions in the occipitotemporal region of the brain, where the ventral stream ends, found it difficult to identify and describe objects while they could move around with seeming ease. Also, patients with lesions in the posterior parietal region, where the dorsal stream ends, were unable to manoeuvre accurately but were able to recognize objects. Monkeys with lesions in inferotemporal regions, having poor visual recognition, were found to be adept at reaching out for moving objects, such as catching flies. Such studies have given strong evidence in favour of the two stream hypothesis. DeYoe and Van Essen [19] suggested that the parietal and temporal lobes could both be involved in shape analysis but associated with different computational strategies. Goodale et al. [20] have suggested that not all illusions affect actions such as grasping or reaching. They contend that actions operate in real time and hence, use the metrics of the real world. They also observed that the more skilled the action, the more likely it will be mediated by the left-hemisphere. Earlier studies have shown that target-directed movements with the right hand are more severely impaired following left-hemisphere damage compared to the other case where left-hand movements are affected due to right-hemisphere damage. Their views that not all illusions affect action are supported by work done by Milner and Dyde [21] who observed differences in action judgements when users reached out for a rod and frame illusion (RFI) and while they reached out for a simultaneous tilt illusion (STI). They found that users twisted their wrists corresponding to the angle perceived in the illusion when reaching out for the STI but in the case of RFI, their action was not fooled although perception was affected. They attribute this to the two illusions being processed in two different areas of the brain. While the STI is processed early in the visual stream, the RFI is processed much deeper in the ventral stream. Thus, while the experiments with the STI shows association between perception and action, the experiments with the RFI show dissociation between the two.

Hughes et al. [22] performed a different experiment to study the dissociation between perception and action. Their experiments required participants to complete a standard line bisection task and a rod bisection task. When asked to locate the centre of the rod, participants showed a rightward bias but while asked to pick up the rods by the centre, their action judgements did not show any bias. Rizzolatti and Matelli [23] concluded that perception and action depend on the activity of the same area in the brain. It is generally believed that perception precedes action. However, they suggested that prior motor knowledge of the external world and actions is used for perception and later action. Glover and Dixon [6] proposed a "planning and control" model where they claimed that visual illusions affect grasping only in the initial stages of movement where the initial movement is planned using visual cues. Subsequently, during reach, correction or control is made irrespective of the presence of visual cues. The effect of the illusion on perception was greater than the effect on action in the last 60% of the reach. As users approached the centre circle, they corrected their grip apertures to its diameter. This model was later contested by Danckert et al. [5] where they found that the maximum grip aperture was unaffected by the size-contrast illusion. Their experimental results showed that the illusion did not affect grasping movement even during the early stages of grabbing and hence their argument against a planning and control model. They argue that differences in the maximum grip aperture seen in the traditional Ebbinghaus display was not due to size-contrast illusion but rather the visuomotor system's attempt to avoid obstacles. They also suggested that visual context can influence action performed on visual illusions through means that are not perceptual.

However, the two-stream theory has been heavily contested by others including Pavani et al. [24] and Franz et al. [25] who propose the existence of a single processing stream. They argue that previous experiments were not conducted in similar environments. Pavani et al. noted that "while perception was subjected to the simultaneous influence of the large and small circles displays, in the grasping task only the annulus of circles surrounding the target object was influential" [24]. To control for what they considered to be flaws in earlier tests, the inadvertent use of different stimuli employed for the perception and action tasks, they designed a new experiment in which the stimuli were more similar. In addition to using the normal Ebbinghaus illusion to test users, Pavani et al. used a "neutral condition" where the central circles were surrounded by circles of the same size thus cancelling the effect of the illusion. For the perception task, they did not use the entire Ebbinghaus illusion. Instead, users were asked to match the centre circle from one half of the illusion to another circle randomly choosen from a set. In this way, during both perception and action, users concentrate on only one half of the illusion. From their experiments, they found that during perception, circle size was overestimated by 0.2 mm when surrounded by smaller circles and there was an underestimation of 0.5 mm when the central circle was surrounded by larger circles. During action tasks, their results showed an overestimation of 0.2 mm in the small surrounding circles condition and an underestimation of 0.8 mm in the large surrounding circles condition. Both underestimation and overestimation were calculated relative to the grip apertures during neutral conditions. In the neutral condition, participants overestimated the size of the centre circle by 0.1 mm during perception and by 0.2 mm during action. In summary, the magnitude of the illusion determined by the large-circles array was double that caused by the small-circles array. Their findings and those of Franz et al. [25] suggest that action is dependent on perception and hence support the one-stream hypothesis.

1.4 Thesis Outline

A brief idea about the working of the visual system was explained. We also looked into the two-stream hypothesis and the contesting one-stream hypothesis and previous research in the field. We also saw how illusions affect perception and how they are used to study the two-stream hypothesis. In particular, we looked into the Ebbinghaus illusion which is one of the more popular illusions used for studying the two-stream hypothesis. Our goal was then described as testing the effects of perception and action in a virtual world, whether the two-stream hypothesis effect is maintained in it and also the differences in perception and action in a VR environment compared to the real world.

The remainder of this thesis is organized as follows. The hardware and software used for our experiments, and a justification for their choice are provided in Chapter 2. Our experiments are described in Chapter 3 and their results described in Chapter 4. Finally, conclusions and future work obtained from our experiments are presented in Chapter 5.

Chapter 2

Hardware and Software

In this chapter, we describe the hardware components and the corresponding software used for running our experiments. Our requirements included a Head-Mounted Display, cameras to test the stereo display of the HMD and motion capture cameras and trackers for tracking the movement of the fingers.

2.1 Hardware Components

The following section explains the various hardware components and the rationale behind the choices.

2.1.1 Head-Mounted Displays

For our experiments, we use an eMagin Z800 3D visor, seen in Figure 2.1, for displaying the Ebbinghaus illusion. The eMagin Z800 consists of 2 OLEDs, each 15.494 cm in length and 8.763 cm in width and with a depth of 2.921 cm providing a stereo display with a resolution of 800×600 and a 40 degree diagonal FOV. The Z800 allows for 360 degrees headtracking horizontally and more than 60 degrees vertically. It is powered by USB or a 5V DC regulated power supply. The EMagin has an RGB Signal Input (PC D-Sub) 24 bit per pixel color. The Z800 also has adjustable interpupillary distance and tilt adjustment which prove to be very useful for performing a part of our experiment described later.

Figure 2.1 eMagin Z800 3D visor

The headtracking capability of the Z800 allows the Ebbinghaus illusion to be displayed in a perspective view. As users move their head, the illusion is moved in the opposite direction to give them a perspective view of the illusion. The position of the HMD is obtained as filtered quaternions which are then converted into Euler angles or axis-angle representations to move the spheres. However, we decided against using the Z800's headtracking as it was found to be noisy and sensitive to the slightest movement which led to objects moving in the display with the slightest twitch. Given that we required the use of a motion capture system to track the position of the fingers, the tracking system of the HMD was not necessary. Finally, our experiments do not involve any change in the position of the illusion. The perspective view was used only for the users to get used to wearing the HMD and movement of objects displayed within the HMD.

2.1.2 Cameras

To test the stereo display of the HMD, we used two Point Grey Flea 2 Model FL2-08S2C cameras each with a maximum resolution of 1032×776 pixels with color in YUV or RGB format. Figure 2.2 shows a Point Grey Flea-2 camera. The maximum bandwidth of the camera is limited by the Firewire B bus. The output from each camera was scaled to a 800×600 resolution before being fed into the HMD.

Figure 2.2 Point Grey Flea-2 Camera

Each camera has a field of view of approximately 1.5×1 meters using a 8 mm monofocal lens. The cameras are placed above the eyes at the interpupillary distance. Output from each camera is fed into the individual displays of the HMD. With the cameras placed at the interpupillary distance, the HMD displays a proper 3D image. However, when the cameras are not properly placed, the scene appears perceptually distorted.

2.1.3 Motion Capture Cameras

In order to track the finger movement during the tasks performed in our experiment, we used NaturalPoint Optitrack motion capture cameras shown in Figure 2.3. The cameras had a focal length of 4.5 mm with a 46 degree horizontal FOV and a frame rate of 100 FPS. The cameras had an imager resolution of 640×480, sub-millimeter accuracy and a latency of 10 ms.

All the cameras in the setup were controlled by a global shutter. The cameras used a USB2.0 cable for data transmission, multiple camera syncing and power. Standard 5V DC can also be used to power the cameras.

2.1.4 Vicon

In addition to testing only the Ebbinghaus illusion in a virtual world, we tested the illusion when surrounded by other objects at various depths. For this purpose, we used a Vicon motion capture system. Six Vicon cameras were placed overhead in a cave environment. Figure 2.4 shows a Vicon camera. Given that the cameras were placed overhead, the Vicon system tracking was very effective as the user's hand movement was not occluded

Figure 2.3 NaturalPoint Optitrack Camera

by any means. With the Optitrack system, we had to take a conscious effort to ensure that the cameras were placed such that the user's hands were not occluded.

The Vicon system also has an inbuilt software component that transmits bundled OSC messages. For the Optitrack system, we had to create our own code to obtain position data and transmit it as OSC messages.

Figure 2.4 Vicon Motion capture camera

2.2 Software Components

The HMD libraries were run under a Linux Ubuntu 10.10 system with an nVidia NV41GL [Quadro FX 400] graphics card. We used the libz800[1] library to run the HMD.

1. http://code.google.com/p/libz800/

A Windows system is required to run the software for both Vicon and NaturalPoint's Optitrack. x, y and z coordinates of the markers corresponding to the finger and thumb are sent from one system to the other. For the headtracking, data is sent as quaternions which are then converted into Euler Angles or Axis-Angle representations before being used to rotate the image displayed. Conversion from quaternion to Euler Angles and Axis-Angle representations is given in Appendix B. To obtain data from the motion capture software, we use the NatNet libraries provided by NaturalPoint and obtain the data as float values. These are then sent across to the machine running the HMD to move the virtual fingers which are used for grabbing the Ebbinghaus illusion in the virtual world.

2.2.1 Virtual Fingers

The virtual fingers and the Ebbinghaus illusion were recreated using OpenGL and C++. We decided on using green cylinders to represent the fingers with one slightly thicker than the other thus representing the thumb and the index finger. We tested models created using Art of Illusion, Blender and Daz 3D to represent the fingers. These were then converted into object files to be displayed using OpenGL. However, we found that the processing overload with models that used a lot of datapoints to represent the data was very high because of which there was a significant delay between finger movement in the real world and that in the virtual world. We used two sets of markers, one for the thumb and the other for the index finger. The motion capture cameras require that each set of markers contain a minimum of two markers. Thus we used two markers for one and three for the other to avoid duplication of marker sets. Because of this constraint, we could not give complete freedom for all the phalanx joints of the two fingers. While it is possible to create complicated models using OpenGL, given that we could not give complete freedom for the finger joints, creating a complex model does not offer much of a benefit over a simple cylindrical model and hence the decision to use the two green cylinders.

2.2.2 Open Sound Control

In order to reflect the movement of the fingers and the head, their positional data has to be transferred from the motion capture system to the HMD system. To do this, we use the Liblo version of Open Sound Control (OSC). Open Sound Control is a protocol for communication among computers, sound synthesizers, and other multimedia devices that is optimized for modern networking technology. One of the primary features that OSC offers is the use of message bundling for those messages whose effects must occur simultaneously. For our purpose, the positional data of the finger and thumb (x,y and z coordinate values for both) must be used simultaneously. The liblo implementation uses TCP and/or UDP for data transport and can be used across platforms. Additionally, it has a high speed transfer with over 100 Hz packet rate. Given that it is an implementation of OSC, it also offers bundle support and timetag support.

2.3 Overall Environment

We covered the hardware and software used in our experiments in this section. While we used the HMD for the virtual environment, we used a board with caps for creating the Ebbinghaus illusion in the real world. Figure 2.5 shows the working of the full system in the virtual world. The movement of the fingers are tracked using motion capture cameras which then send data from a Windows system to a Linux machine running the HMD libraries.

This data is sent using the liblo implementation of OSC.

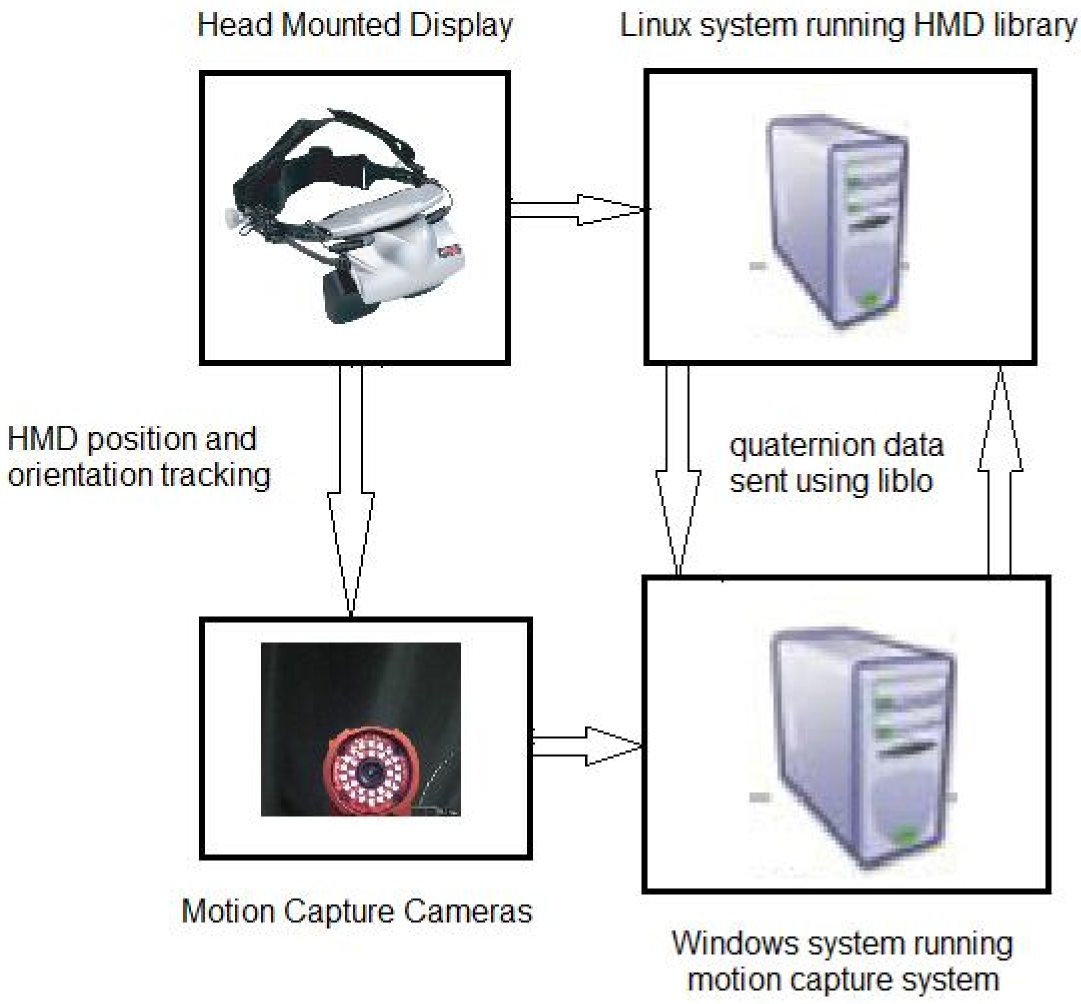

Figure 2.5 System architecture

Chapter 3

Experiments

The two-stream hypothesis states that there is a dissociation between perception and action. Thus, in order to test the hypothesis, we had to test how users perceive an illusion and how they act upon it. During perception, users were asked to give an estimate of the diameter of the two centre circles in the illusion. During action, users were asked to grab the illusion. In the real world, they used their fingers while in the virtual world, two green cylinders represented their fingers. As mentioned earlier, we wanted to test the two-stream hypothesis in a virtual world. In addition to that, we wanted to see the differences while grabbing in a real world and in a virtual world. With the lack of haptic feedback in a virtual world, it would be quite interesting to see how the brain processes action tasks in such an environment and how action tasks differ from that in the real world.

3.1 Experimental Setup and Methods

Twenty participants were used for the study. The age of the participants ranged from 22 to 39. There were a total of 13 male subjects and 7 female subjects. All of the participants were right handed. The experimental procedure was divided into a real world phase and a virtual world phase. The user documents are attached in Appendix A. Each of these phases was further subdivided into a perception task and an action task. The tasks involved estimating the diameter of a sphere and grabbing it using the index finger and thumb. Our tracking setup consisted of six 4.5 mm lens Optitrack cameras. Three cameras were placed on either side of the user at a horizontal distance of approximately one metre. The cameras were placed at a height of around 1.5 m from ground level. The cameras were placed such that they faced both the illusion and the user's hands at the same time. This allowed for tracking to be present during grabbing. Two sets of reflective markers were placed on the user's right hand index finger and thumb respectively. These were then tracked as two separate objects by the cameras. The position of these two objects was taken directly as their x, y and z positions with respect to the origin. One additional trackable object was placed on top of the HMD to track it and hence allow the user to have a perspective view. As mentioned earlier, the perspective view was used only for the user to get used to the HMD and movement of objects within the HMD. For this purpose, we used a single wired sphere rather than the Ebbinghaus illusion itself.

3.2 Real World Phase

In the real world phase, users saw the Ebbinghaus illusion mounted on a table. The illusion was placed around 0.5 m away from the user. The illusion, shown in Figure 3.1 was created using table leg caps and felt pads fixed to a wooden board which was placed vertically on a table. The central cap in the illusion was 1" in diameter. The bigger surrounding pads were 1 1/2" in diameter and the smaller surrounding pads were 1/2" in diameter. The sizes were chosen to ensure that there was enough size contrast for the user to be fooled. The central caps were much thicker than the surrounding caps to allow for grabbing. We shall refer to both the caps and pads as spheres for the purposes of explaining the experimental procedure.

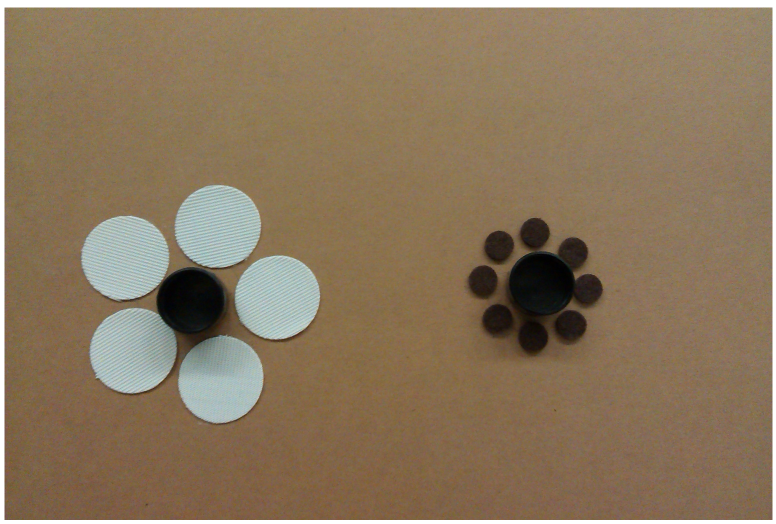

Figure 3.1 Ebbinghaus recreated in the real world

During the **perception task**, users were asked to spread their thumb and index finger to approximate the diameters of both the central spheres, one by one. The two approximated diameters, which are the distance between the thumb and index finger for each sphere, were measured as r_{p1} and r_{p2} where r stands for real and p for perception. r_{p1} corresponds to the sphere surrounded by smaller spheres while r_{p2} corresponds to the sphere surrounded by larger spheres. As will be explained later, users could not occlude the illusion during perception in the virtual world. To preserve this effect in the real world, we ensured that the users had their hands to the side of the illusion and not directly in front of it as seen in Figure 3.2.

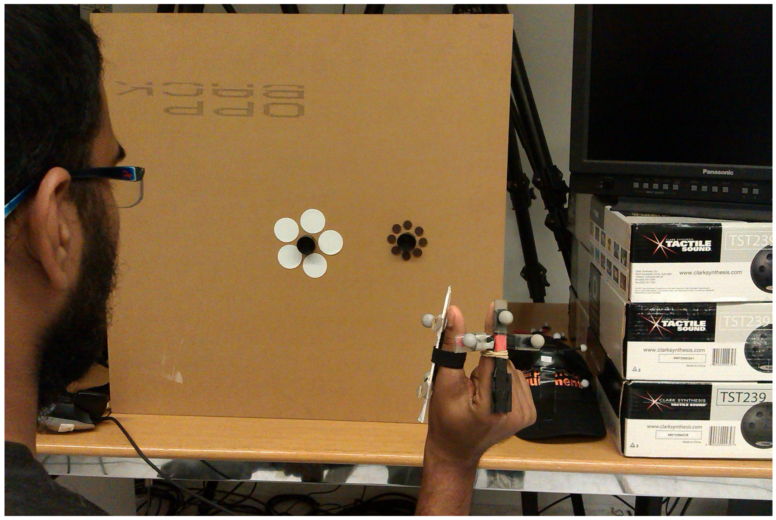

Figure 3.2 User approximating a circle in the real world

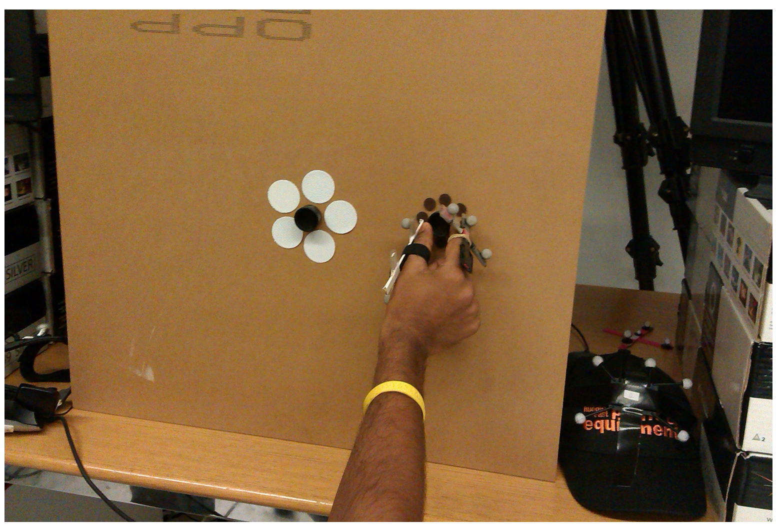

Figure 3.3 User grabbing a circle in the real world

During the **action task**, users were asked to grab the central spheres, one by one, using their thumb and index fingers. This is shown in Figure 3.3 The grabbing apertures, which are the distance between the thumb and index finger for each of the spheres, were measured as r_{a1} and r_{a2} where a stands for action. Users were asked to bring their fingers to a neutral position before grabbing or approximating the size of each of the spheres. This was to ensure that there were no residual effects of the previous approximation or grabbing. The first part tests the user's perception while the second part tests the user's vision based action. Also, during this phase, users did not wear a HMD and were able to see their hands at all times (closed loop condition).

3.3 Virtual World Phase

Before users were shown the Ebbinghaus illusion in the virtual world, they were asked to get accustomed to moving their fingers in the virtual world. Two green cylinders corresponding to the user's finger and thumb were shown in the display. Users were able to translate the cylinders in the x,y and z directions. Rotation was however restricted as noise affected the movement of the cylinders in the virtual world. Once the user was accustomed to moving the cylinders in the virtual world, we proceeded with conducting the actual experiment in the virtual world.

The experimental procedure for the virtual world phase was very similar to the real world phase. For the **perception task**, users were shown the Ebbinghaus illusion in the HMD and asked to spread their index finger and thumb to approximate the diameters, and hence the radii, of the two centre spheres in the illusion. Users were able to see their hand in the real world during this phase of the experiment to ensure that they had an idea of how much they had approximated the diameter of the centre sphere. Users had a choice of viewing their hands from the corner or bottom of their eye or opening the flap of the Emagin HMD in order to see their hands as shown in Figure 3.4. During this phase, users

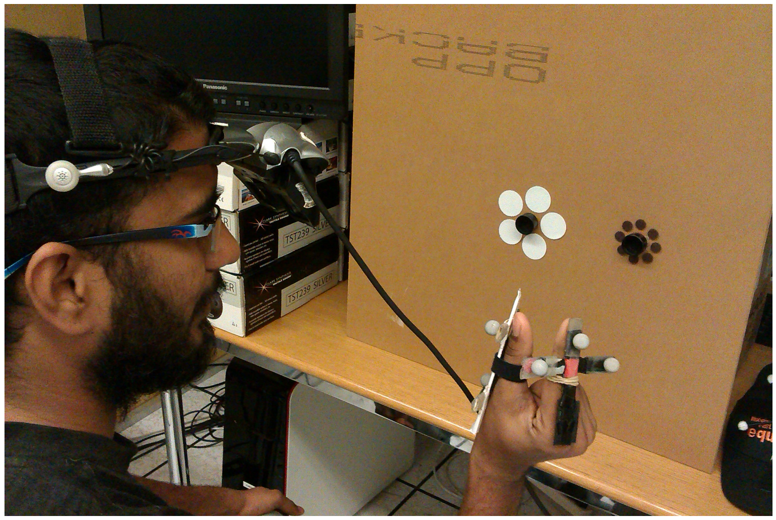

Figure 3.4 User approximating a circle in the virtual world

could not occlude the illusion using their fingers. The distance between the index finger and thumb was measured as v_{p1} and v_{p2} for each of the spheres. v_{p1} corresponds to the sphere surrounded by smaller spheres while v_{p2} corresponds to the sphere surrounded by larger spheres. v denotes that the task was performed in a virtual world.

For the **action task**, as seen in Figure 3.5 subjects were asked to "grab" the centre spheres, with visual feedback provided to indicate the positions of their thumb and forefinger. Two green cylinders were used to represent the fingers. This can be seen in Figure 3.6 The finger separations (grip aperture or grabbing aperture) were measured as v_{a1} and v_{a2}.

Similar to the real world, it was ensured that users brought their fingers to a neutral position between each task. In both the real and virtual world phases of the experiment, the system was stopped as soon as users reported that they had approximated or grabbed the illusion. The measurements of the apertures were averaged over time as the user approximates or grabs the illusion. Maximum grip apertures were also measured.

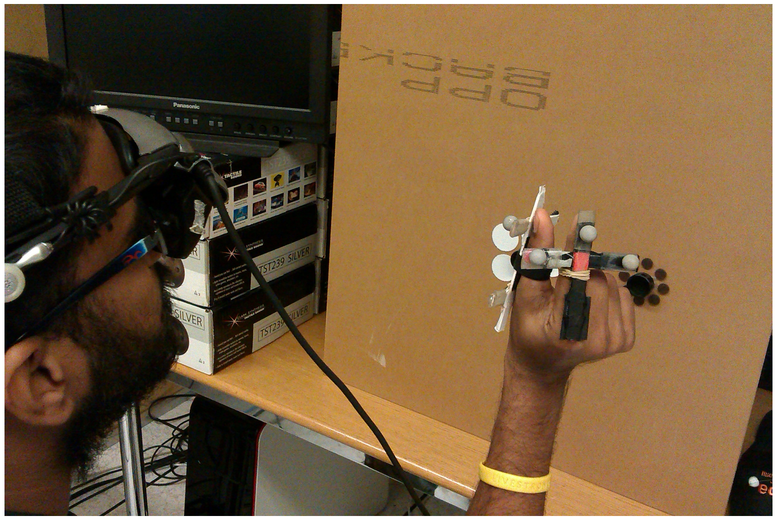

Figure 3.5 User grabbing a circle in a virtual world

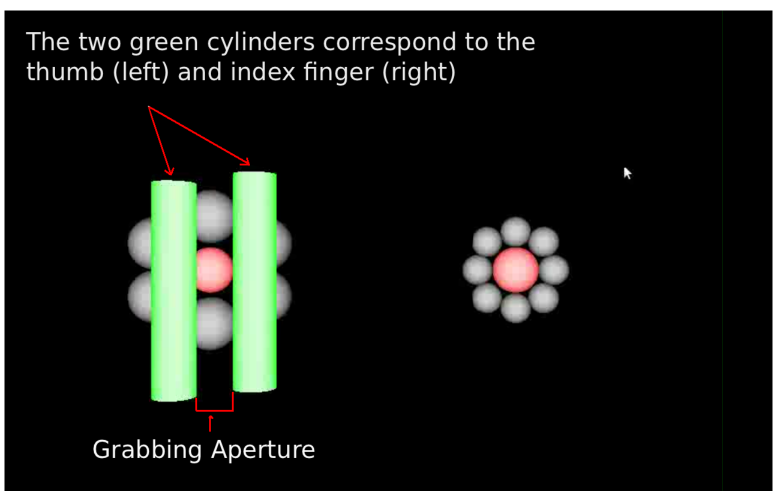

Figure 3.6 User's view of grabbing a circle in the virtual world

3.4 Depth Testing

From our results, explained later, we found that around half the number of participants grossly underestimated depth in the virtual world. We decided to test the illusion along with a few depth cues to see how it affected users grabbing spheres in the virtual world. We used a simple wired sphere and a teapot placed on either side of the illusion along the depth axis. We also had a depth grid covering three walls. In addition to this, we introduced colour changes to help participants estimate depth. As seen in Figure 3.7, users had to grab the illusion when the cylinders turned red notifying them that they had reached the correct depth plane. An error of +/-1 was allowed in the depth axis as getting both virtual fingers in the same depth plane was an extremely hard task. Thus we use three different depth cues namely, depth grid, colour changes and objects placed on depth axis to help users gauge depth. While the other depth cues are used for gauging coarse distance, the colour change is primarily used for finer adjustments such as when the user is around the depth plane. We then tested users' perception and action similar to testing them in virtual world as described in Section 3.3.

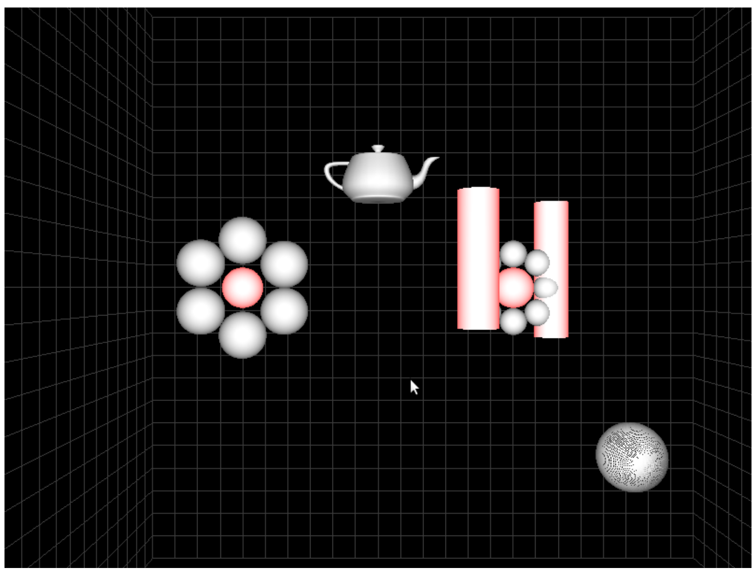

Figure 3.7 User's view of grabbing a circle in the virtual world. Additional depth cues can be seen

Chapter 4

Results and Conclusions

We first describe the possible outcomes of the experiments and their interpretations. Later we look into the results and how they conform to the outcomes detailed below.

4.1 Possible Outcomes

For our studies, the absolute differences $r_{pd} = |r_{p1} - r_{p2}|$ and $r_{ad} = |r_{a1} - r_{a2}|$ and similarly $v_{pd} = |v_{p1} - v_{p2}|$ and $v_{ad} = |v_{a2} - v_{a2}|$ are the metrics necessary for analysis. r_{pd} is a measure of how much perception has been fooled in the real environment. Given that it is the difference between the approximated diameters of each of the central spheres, the higher r_{pd} is, the bigger the perceived difference in radii of the two spheres. Thus, a high r_{pd} indicates that the user has been fooled by the illusion. When r_{pd} is significantly large, it implies that while approximating the two spheres during the perception task, the user has approximated one sphere, r_{p1} bigger than the other, r_{p2}. Conversely, a low r_{pd} would mean that perception has not been fooled. However, previous research in general has shown that the possibility of a low r_{pd} is minimal. In other terms, there is a very less chance that perception is not fooled by the illusion and one of the major reasons for perception not being fooled could be previous knowledge of the illusion. From the action tasks results, r_{ad} is a measure of how much action has been fooled in the real environment. The higher the difference between grabbing apertures (r_{ad}), the more action has been fooled by the illusion. If r_{ad} is small, the user has scaled his grip aperture such that the user grabs both the central spheres correctly which in turn would mean that while grabbing the spheres, the user's motor system 'believes' that both the spheres are of the same size. However, the user would have still perceived the spheres' diameters incorrectly. v_{pd} and v_{ad} are defined similarly in the virtual world. According to the two-stream hypothesis, perception and action are processed in two different streams and as such illusions fool perception but not action. Thus when the difference in perceptual approximations (r_{pd}) is significantly larger than the difference in grabbing (action) apertures (r_{ad}), the results lend significant weight to the two-stream hypothesis. However, if the two differences approximate each other, then it adds weight to the one-stream hypothesis. This can be drawn from the fact that both perception and action have been fooled by the illusion. Similar conclusions can be drawn for the virtual environment scenario. Table 4.1 summarizes the above points.

Table 4.1 Summary of measurements used and the inferences drawn

<table>
<tr><td>$v_{pd} > v_{ad}$</td><td rowspan="2">Virtual World</td><td>Perception is fooled but action is not</td></tr>
<tr><td>$v_{pd} = v_{ad}$</td><td>Both perception and action are fooled</td></tr>
<tr><td>$r_{pd} > r_{ad}$</td><td rowspan="2">Real World</td><td>Perception is fooled but action is not</td></tr>
<tr><td>$r_{pd} = r_{ad}$</td><td>Both perception and action are fooled</td></tr>
<tr><td>$r_{ad} \not\approx v_{ad}$</td><td colspan="2">Indicates integration of proprioceptive and visual feedback</td></tr>
<tr><td>$r_{ad} \approx v_{ad}$</td><td colspan="2">Indicates similarity in action mechanisms in both real and virtual worlds</td></tr>
</table>

If v_{pd} is large, this would mean that the user's perception is fooled by the illusion similar to the real world scenario. The question, as mentioned earlier, is how users grab the central spheres in the virtual world and how the indirect relation between proprioceptive and visual feedback affects grabbing apertures of the users. In the virtual world, users will only rely on visual feedback to guide their movements (as would be expected in a virtual environment) and as such, we should observe a gradual refinement of the grip aperture to fit the size of the central sphere during the course of the action task. If we observe something characteristically different between how this task is performed in the real world versus the virtual environment, then that would to some degree, confirm that the process of vision based action relies on the integration between proprioceptive and visual modalities in the real world. However, if the real and virtual world action measurements are similar, i.e, if this task is performed similarly in both real and virtual environments, with one having an indirect feedback system and the other having a direct feedback system; it should raise questions about why action is similar in both virtual and real environments even with a lack of haptic feedback in the former.

4.2 Analysis and Discussion

We had earlier discussed about the results and their significances. Summarising it, if r_{pd} is found to be significantly greater than r_{ad}, then it would mean that our results would tend towards the two-stream hypothesis. However, if both perception and action measurements are close to each other or if action grip apertures are greater than perception apertures, it would mean that our results tend towards the one-stream hypothesis.

From the obtained data, we do not know the exact difference value from which we can determine if perception or action has been fooled. Consequently, we cannot determine which hypothesis our results tend to without knowing if perception and action has been fooled. A solution to this is to use the distributions of perception and action with the equality between the two determining the hypothesis. We performed a two-sample Kolgomorov Smirnov test on the obtained data. The K-S test is a standard non-parametric test used to calculate the equality between distributions. One of the advantages of the K-S test is that it does not take into consideration the distribution of the data. This can also be considered as a disadvantage as there are more powerful tests that can be performed on data if the distribution is known before hand. Given that we do not know the distribution of the data, the K-S test is very useful for our analysis.

The K-S test on perception and action data in the real world, using the average grip apertures show that both the distributions have a difference of 0.4500 with a p-value [1] of 0.023. The K-S test rejects the null hypothesis that both sets of data are from similar distributions provided the p-value is lesser than a chosen α of 0.05. This chosen value of α is arbitrary. Table 4.2 shows the p-values for the K-S tests on the distributions of perception and action in real and virtual worlds.

This means that in the real world, perception and action are from different distributions. From the cumulative fraction plot shown in Figure 4.1, it can be seen that for mostly any value on the x-axis, i.e., grip aperture values, the fraction of perception values strictly lesser than the given grip aperture is less than the fraction of action values lesser than the

1. p-value is defined as the probability, if the test statistic really were distributed as it would be under the null hypothesis, of observing a test statistic [as extreme as, or more extreme than] the one actually observed.

Table 4.2 P-values from the Kolgomorov Smirnov test using average grip apertures

Environment	K-S test between	Corresponding P-value	Difference value
Real	Perception and action	0.023	0.4500
Virtual	Perception and action	0.771	0.2000
Real and Virtual	Action in real and virtual world	0.008	0.5000

given grip aperture. This in turn means that for the most part, the perception grip aperture differences are much larger than the action grip apertures. Thus we can see that results tend towards the two-stream hypothesis in the real world.

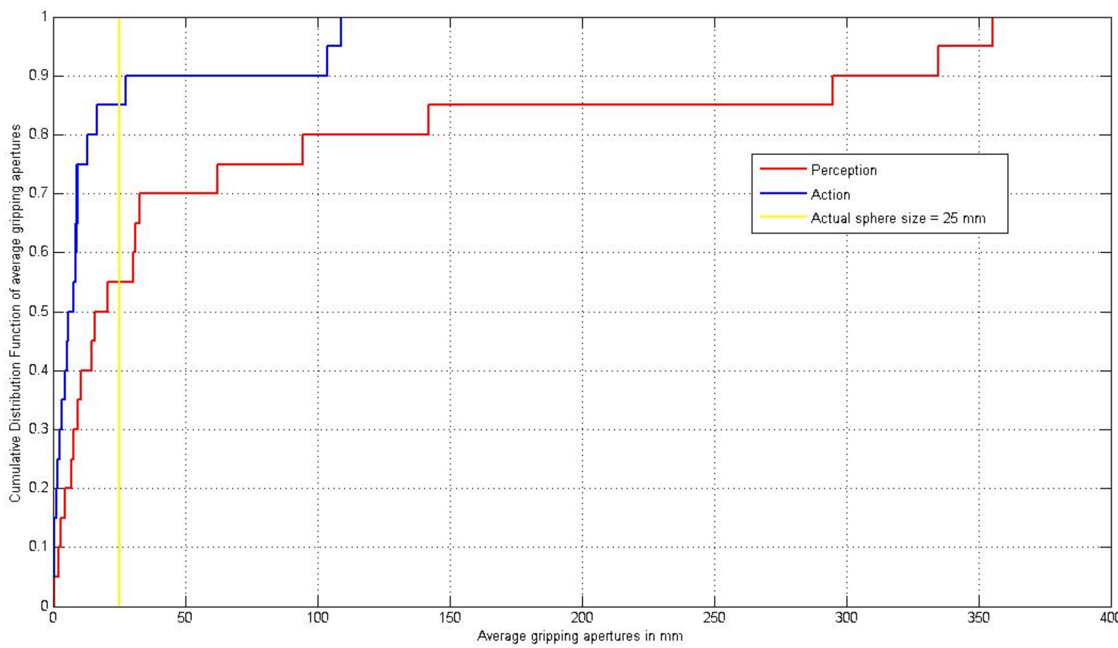

Figure 4.1 Cumulative Fraction plot of average gripping apertures in Real world

In the virtual world, using the average grip apertures, we see that the K-S test between perception and action values shows a distribution difference of 0.2000 with a p-value of

0.771. The cumulative fraction plot seen in Figure 4.2 shows that both the perception and action grip apertures are similar to each other. It can be seen that both plots are intertwined and hence the data sets are from similar distributions. The results can be said to tend towards the one-stream hypothesis in the virtual world. This can be explained by the fact that in the virtual world we do not have a haptic feedback mechanism. The only feedback that is present is visual and hence action is a result of perception in virtual environments. It was also seen that in the virtual world, participants underestimated depth compared to real world. Most of the participants assumed that they had grabbed the illusion even when the illusion was located much farther on the depth axis.

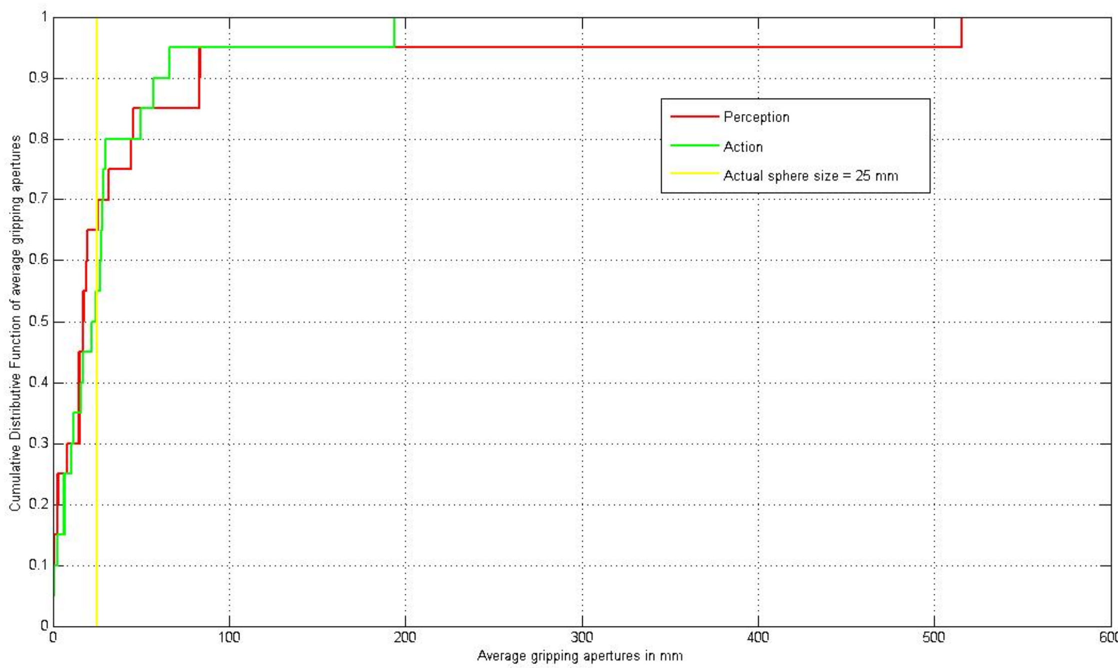

Figure 4.2 Cumulative Fraction plot of average gripping apertures in Virtual world

We also measured the distribution difference in action values across both real and virtual worlds, where we found that the K-S test gives a distribution difference of 0.5000 with a p-value of 0.008. We can again see that the action measurements across both worlds come from different distributions. Given that the hypothesis varies across real and virtual worlds, it should follow that action results are dissimilar between real and virtual worlds. From Figure 4.3, it can be seen that the grip apertures during action in the virtual world are higher than the grip apertures during action in the real world. Given that users are fooled in the virtual world even during action, their grip aperture differences in action are higher than that in the real world where they are not fooled.

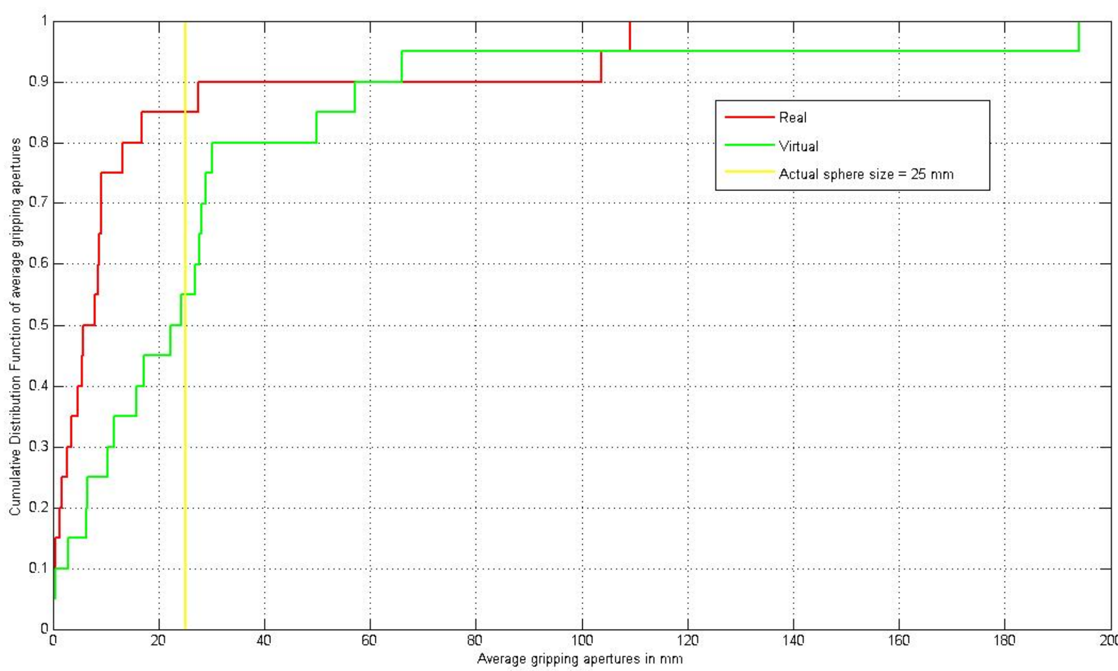

Figure 4.3 Cumulative Fraction plot of average gripping apertures during action in both worlds

In general, we see that perception grip apertures are higher than action grip apertures in the real world. Also, action grip apertures in the virtual world are higher than action grip apertures in the real world. Finally, in the virtual world, perception and action grip apertures are very similar. It can hence be said that in the real world, the brain seems to operate using two streams to process perception and vision based action. In the virtual world however, the brain uses visual feedback as its only cue to process both perception and action. We also note that the lack of proprioceptive feedback in the virtual world results in action in the virtual world being processed differently compared to action in the real world.

Apart from using average grip aperture, we had measured maximum grip apertures too. During approximation or grabbing the sphere, users might have moved their fingers to a more comfortable position thus leading to stray values. This can result in spurious maximum apertures. To avoid this, we eliminated outliers before using the maximum grip apertures to find the distribution difference between perception and action. Our results were found to be very similar to the results obtained using average grip apertures. As can be seen from Table 4.3, perception and action in the real world come from different distributions while in the virtual world, they are from the same distribution. Action across both worlds are from different distributions.

Table 4.3 P-values from the Kolgomorov Smirnov test using maximum grip apertures

Environment	K-S test between	Corresponding P-value	Difference value
Real	Perception and action	0.005	0.6643
Virtual	Perception and action	0.446	0.4457
Real and Virtual	Action in real and virtual world	0.017	0.5490

Figures 4.4, 4.5 and 4.6 show the cumulative fraction plots of the illusion in the real world, the virtual world and action across both worlds respectively. It can be seen that the

plots are very similar to the plots obtained using average grip apertures.

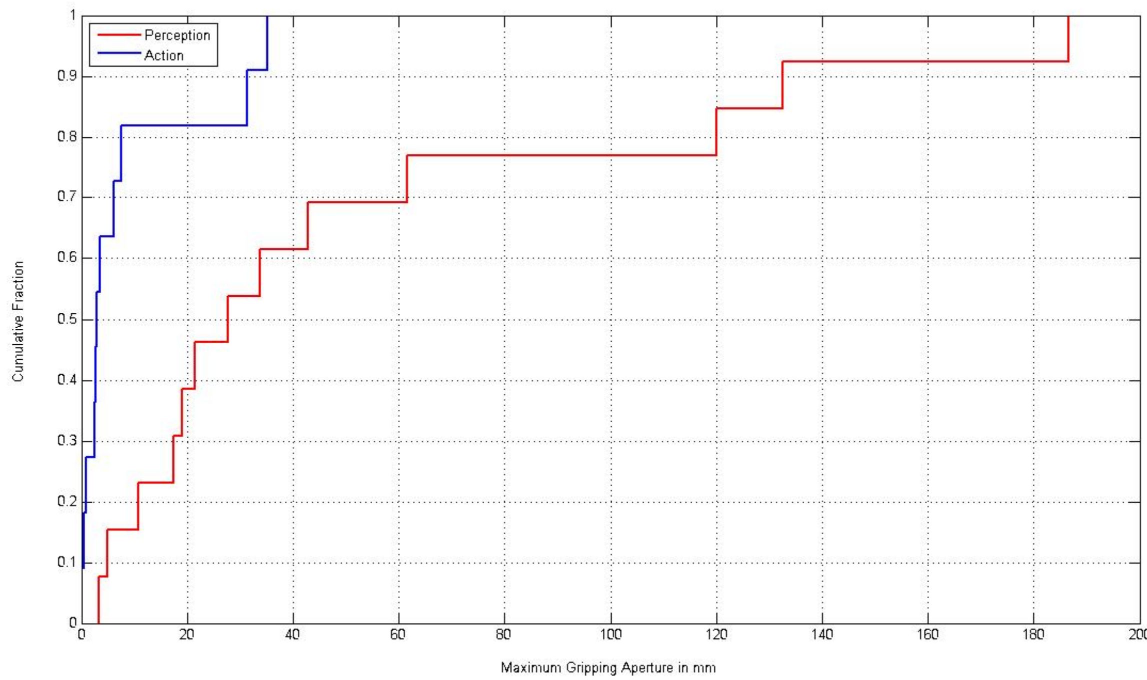

Figure 4.4 Cumulative Fraction plot of maximum gripping apertures in real world

As mentioned earlier, users grossly underestimated depth in the virtual world. The average grabbing depth is shown in Figure 4.7. The sphere was located at around 10 OpenGL units from the origin along the depth axis. Here, average grabbing depth refers to the averaged depth plane where the users thought that they had grabbed the illusion. It was found that close to half the number of participants underestimated depth in the virtual world by more than 30%.

We later tested a few subjects using depth cues as detailed in Section 3.4. Our main aim was to find out if underestimation of depth affected the gripping aperture differences.

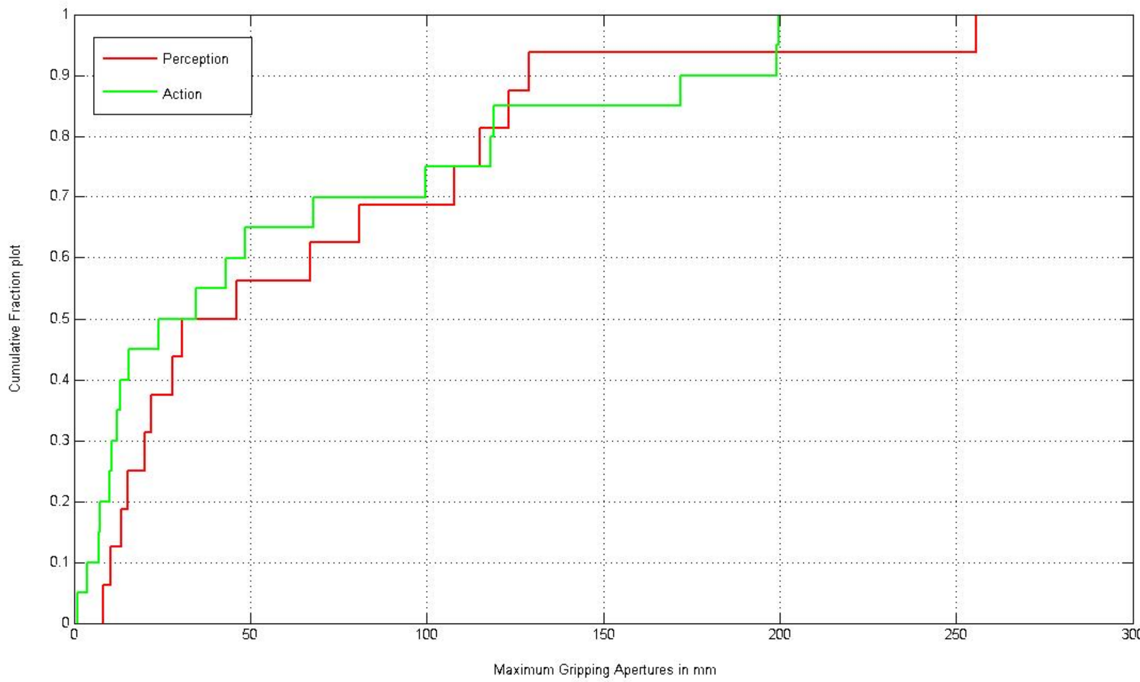

Figure 4.5 Cumulative Fraction plot of maximum gripping apertures in virtual world

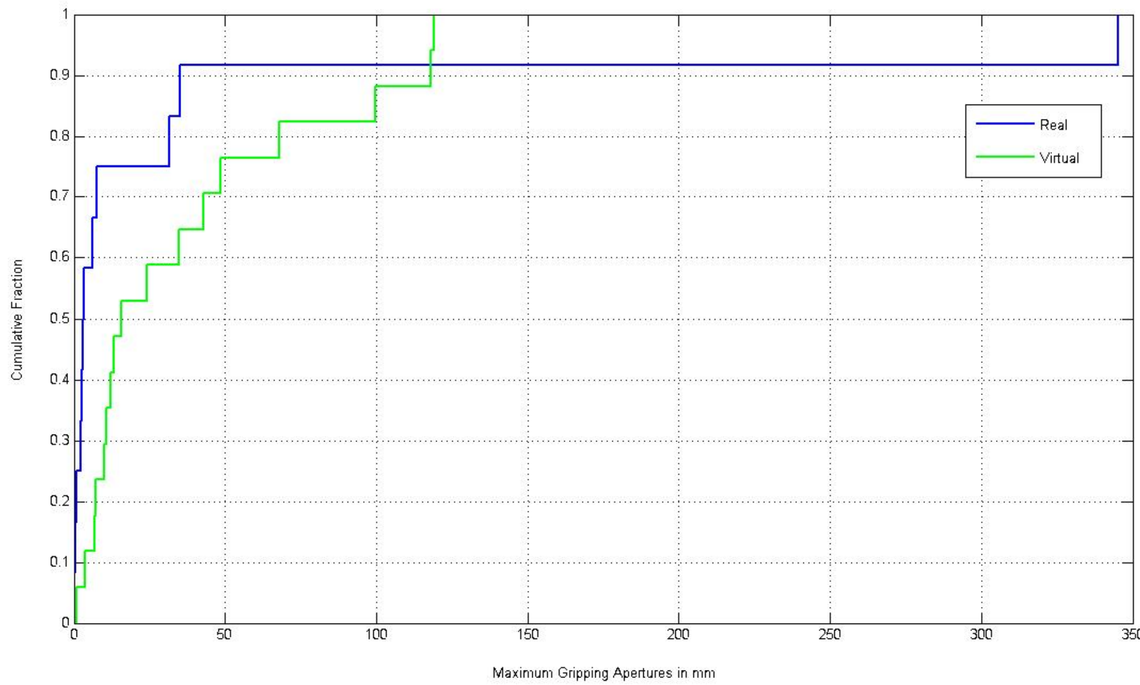

Figure 4.6 Cumulative Fraction plot of maximum gripping apertures during action in both worlds

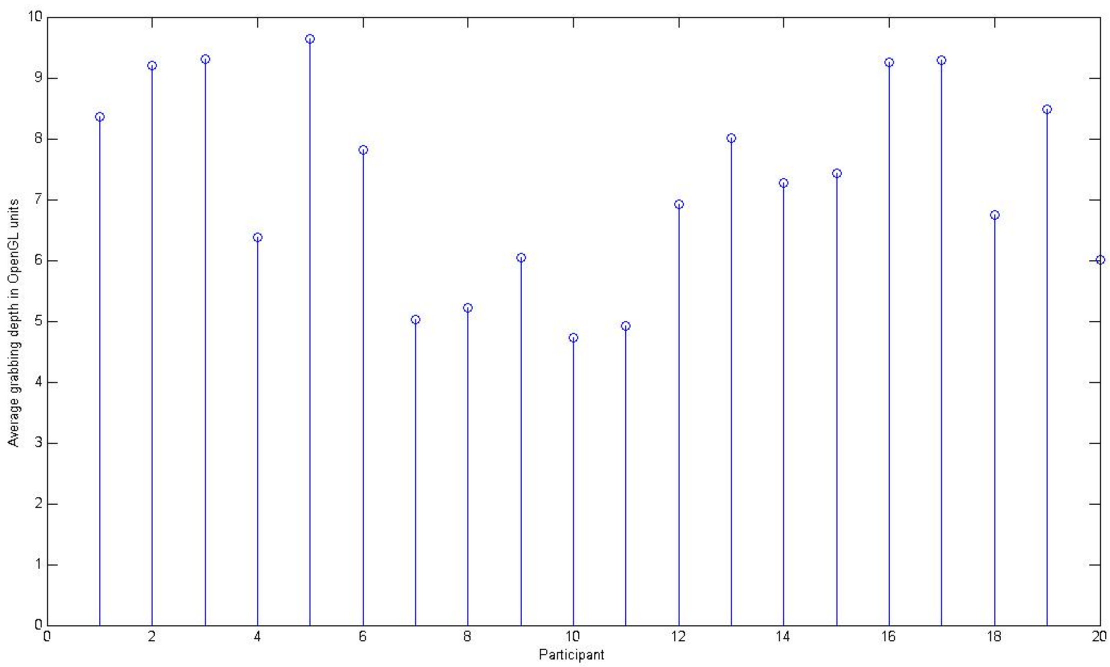

Figure 4.7 Average Grabbing Depth

We found that even with depth cues, the results tended towards the one-stream hypothesis. The K-S test on perception and action with depth cues in the virtual world resulted in a distribution of difference of 0.2222 and a p-value of 0.9575. This means that both distributions are very similar in nature. Figure 4.8 shows the cumulative fraction plot.

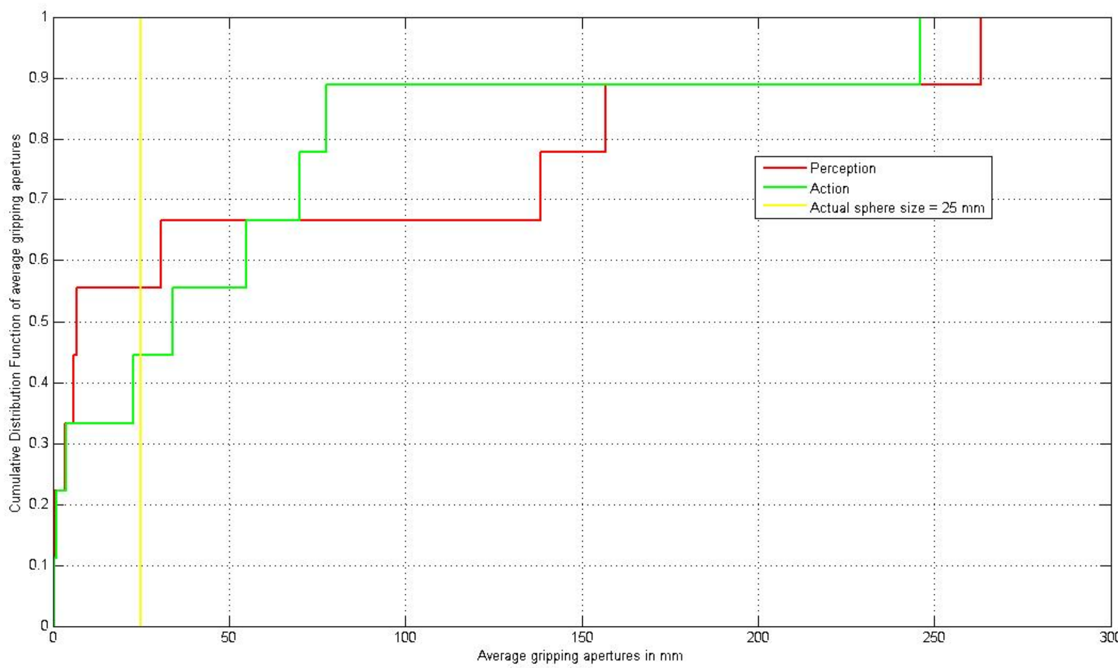

Figure 4.8 Cumulative Fraction plot of maximum gripping apertures during action in both worlds

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Our results show that in the real world the brain processes perception and action in the ventral and dorsal stream respectively while in the virtual world, both perception and action are processed in a single stream, presumably the ventral stream as the feedback is entirely visual.

Our results in the virtual world have also shown that proprioceptive feedback is essential for depth perception while grabbing an object. While stereoscopy allows for the eyes to judge distances of various objects in a scene, to perform an action on the object, some sort of haptic feedback is essential for the brain to process the action. In the real world, when an action is performed, the users' ability to feel the object provides necessary information for the brain to process the action without the need for perceptional cues once the action has commenced. One of the primary examples is a sport like soccer where sometimes players don't necessary look at a ball while they are able to control the ball. However, in the virtual world, it would not be possible to let users know that they are controlling the object unless visual feedback is used. Hence, unless some sort of haptic feedback is provided, users will have to completely rely on perceptional cues in the virtual world to process action.

An interesting point raised by the examiner was that the effect of haptic feedback is felt only at the end of the grabbing motion and not before. Thus, this would suggest that the lack of haptic feedback in the virtual world should not affect the user's motor system from planning the right grip aperture to 'grab' the sphere. However, our results show that, compared to the real world, users' grip apertures differences between perception and action was lesser in the virtual world. During our experiments in the virtual world, users perceptually estimated the size of the central sphere by spreading their fingers apart in the real world while they had to 'grab' using virtual fingers. This mismatch could have an effect on user judgement.

It was also shown that in the virtual world, depth is often underestimated unless a subject receives external cues to judge distances. In our case, we used colour changes in the virtual arms to provide cues to the user about the depth levels. Without these cues, users were quite far away from an object when they were grabbing it while they were under the impression that they had reached the object. This also further proves the need for proprioceptive feedback to gauge depth while performing an action.

5.2 Future Work

While we have drawn many conclusions based on the experiments we had conducted, there is still room for a lot of improvement or changes in the experimental conditions. For instance, in the virtual world, the users had only the illusion shown to them. However, in the real world, along with the illusion, users could also see surrounding objects that were part of the experimental room. The presence of such objects located away from the illusion, in theory, should not affect users' judgement of the sizes of spheres in the illusion.

One of the major limitations in our experiments was the lack of realistic virtual fingers. As had been mentioned earlier, the motion capture system required a minimum of two markers for each object being tracked and additionally these markers should be differently oriented so they are not duplicated. Creating more realistic virtual fingers could improve the users' experience during the experiments. There is also the question of whether more realistic virtual fingers affect grabbing apertures. We had used Vicon and NaturalPoint Optitrack for motion capture. While they were very effective, they had to be calibrated meticulously and ensured that they were not moved after calibration. In addition to this, markers had to be placed effectively to ensure good tracking and non-duplication. A wrist-worn device such as the one developed by Microsoft [1] can be used in place of the entire motion capture system.

The experiments were conducted in a closed loop condition where the users could see the illusion at all times. It would be worthwhile to see the experimental results when performed under an open loop condition where the illusion is blocked once the user commences

1. http://www.engadget.com/2012/10/09/microsoft-research-digits-3d-hand-gesture-tracking/

to either approximate or grab the illusion. However, it has to be noted that previous research has shown that other than the time taken to grab the object, lack of visual feedback does not considerably affect gripping apertures. It would also be interesting to see how the inclusion of an external feedback in the virtual world affects depth gauging and user actions. In our second set of experiments on depth gauging, we had users trained using colour changes to gauge depth. By incorporating a haptic feedback mechanism or an auditory feedback which gives a different sound for grabbing objects at various depths, we can have an external feedback that lets users know if they have grabbed the object. While this would certainly influence depth perception, it could also have an affect on the grabbing apertures of users in the virtual world.

As mentioned in the previous section, there could have been a mismatch in the experimental conditions in the virtual world thus leading to an error in user judgement. This could be changed by conducting the perception experiment in the virtual world by asking users to either select a circle from a set of circles that they feel corresponds to the size of the centre circle or by asking them to approximate the circles using virtual fingers. The other factor is if the top-down knowledge effect that there is no physical circle to be grabbed in the virtual world affects the 'how' system. As suggested by the examiner, this could be tested by conducting another experiment in which the participant would be asked to initiate a grasping behavior and only found out after the motor behavior was initiated whether the grasping was real or virtual.

One major application of our research would be in helping recovering stroke patients. Most stroke patients suffer from motor disabilities and hence undergo therapy to help them get back their motor skills. In the real world, patients stand a risk of tripping over objects during rehabilitation sessions. Virtual worlds help them perform these motor actions while at the same time, avoiding the dangers of real objects present in their paths. By compensating for lack of depth perception using scaled objects in the virtual world or through haptic feedback, patients can be helped to recover from disabilites caused by stroke.

Video games are another area where perception and action are used very often in a virtual world. In most video games, action is performed through buttons on a console. More recently, the X-box Kinect allows for natural input instead of button or console input. With the help of external haptic feedback, users can gauge depths better when playing a game on Kinect.

Appendix A

User Documents

McGill

Research Ethics Board Office
James Administration Bldg, room 429
845 Sherbrooke St West
Montreal, QC H3A 0G4

Tel: (514) 398-6831
Fax: (514) 398-4644
Ethics website:www.mcgill.ca/research/researchers/compliance/human/

Research Ethics Board I
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 23-0612

Project Title: 3D Ebinghaus illusion

Principal Investigator: Prof. Ian Gold **Department:** Philosophy and Psychiatry

Co-Investigators/ Other Researchers: Rajkumar Viswanathan, Alireza Hashemi, Jeremy Cooperstock

This project was reviewed by delegated review.

Rex Brynen

Rex Brynen, Ph.D.
Chair, REB I

Approval Period: _19 July 2012_____ **to** __19 July 2013_____

This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

---

* All research involving human participants requires review on an annual basis. A Request for Renewal form should be submitted 2-3 weeks before the above expiry date.
* When a project has been completed or terminated a Study Closure form must be submitted.
* Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received.

CONSENT FORM

In this study you will be asked to perform a few simple motor tasks while wearing a head-mounted display. Your goal is to reach for one of two circles in a virtual environment. You will also be asked to estimate the size of these circles.

The aim of this project is to investgate whether humans employ different mental representations in manipulating objects under different conditions. This helps in the design of a behavioural paradigm that may subsequently be used in Neuroscience research. Our study may also have important implications for the design of immersive virtual environments.

The investigators are: Rajkumar Visnawathan (Dept. of Electrical Engineering), Alireza Hashemi (Integrated Program in Neuroscience), Ian Gold (ian.gold@mcgill.ca, Dept.s of Psychiatry and Philosophy) and Jeremy Cooperstock (Dept. of Electrical Engineering).

Your participation in this study is completely voluntary and you may choose to stop participating at any time. Your decision to stop participating, or refusal to answer particular questions, will not affect your relationship with the researchers, McGill University, or any other group associated with this project.

You will receive a minimum of 5 dollars in compensation for your participation in this study. The expected approximate time for this research is 20 minutes.

All information you supply during the experiment will be held in confidence and your name will not appear in any report or publication of the research. Your data will be safely stored in secure digital folders which only research staff will have access to. Confidentiality will be provided to the fullest extent possible by law.

If you have any further questions about the research in general or about your role in the study, please feel free to contact Alireza Hashemi (alireza.hashemi@gmail.com).

By signing this form you confirm that you have had the opportunity and the time to ask questions, which have been answered by the research team to your satisfaction.

You understand that your participation in this study is entirely voluntary and that you may withdraw at any time during the experiment, without any consequence.

By signing this document, you agree to participate in this research project.

Questionnaire

Gender : Male ☐ Female ☐

Age :

What do you think of the two black/red spheres shown during the trials.

Have you heard of the Ebbinghaus illusion before? If yes, where?

What other geometrical illusions are you familiar with?

Appendix B

Mathematical Conversions

B.1 Quaternion to Euler Angles

Given a quaternion, q, where

$$q = [q_0 \ q_1 \ q_2 \ q_3]^T \tag{B.1}$$

the Euler angles are obtained as

$$\phi = atan2(\frac{2(q_0q_1 + q_2q_3)}{1 - 2(q_1^2 + q_2^2)}) \tag{B.2}$$

$$\theta = arcsin(2(q_0q_2 - q_3q_1)) \tag{B.3}$$

$$\psi = atan2(\frac{2(q_0q_3 + q_1q_2)}{1 - 2(q_2^2 + q_3^2)}) \tag{B.4}$$

where ϕ θ and ψ are the Euler angles along the x, y and z directions respectively. Also, the quaternion, q, is normalized before conversion.

B.2 Quaternion to Axis Angle Representation

Given a quaternion, q, where

$$q = [q_0\ q_1\ q_2\ q_3]^T \tag{B.5}$$

the axis angle representation is obtained as

$$\theta = 2arccos(q_0) \tag{B.6}$$

$$x = \frac{q_1}{1 - q_0^2} \tag{B.7}$$

$$y = \frac{q_2}{1 - q_0^2} \tag{B.8}$$

$$z = \frac{q_3}{1 - q_0^2} \tag{B.9}$$

where x, y and z form the axis along which the rotation θ occurs.

Bibliography

[1] M. Mishkin, L. G. Ungerleider, and K. A. Macko, "Object vision and spatial vision: two cortical pathways," *Trends in Neurosciences*, vol. 6, pp. 414–417, 1983. 1, 1.3

[2] A. M. Haffenden, K. C. Schiff, and M. A. Goodale, "The dissociation between perception and action in the ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp," *Current Biology*, vol. 11, no. 3, pp. 177–181, 2001. 1.1, 1.1

[3] D. W. Massaro and N. H. Anderson, "Judgmental model of the Ebbinghaus illusion," *Journal of Experimental Psychology*, vol. 89, no. 1, pp. 147–151, 1971. 1.1

[4] T. Jaeger and K. Grosso, "Contour lightness and separation effects in the ebbinghaus illusion," *Perceptual and Motor Skills*, vol. 76, no. 1, pp. 255–258, 1993. 1.1

[5] J. Danckert, N. Sharif, A. Haffenden, K. Schiff, and M. Goodale, "A temporal analysis of grasping in the ebbinghaus illusion: planning versus online control," *Experimental Brain Research*, vol. 144, pp. 275–280, 2002. 1.1, 1.3

[6] S. Glover and P. Dixon, "Dynamic effects of the ebbinghaus illusion in grasping: Support for a planning/control model of action," *Attention, Perception and Psychophysics*, vol. 64, pp. 266–278, 2002. 1.1, 1.3

[7] M. T. Bolas, "Human factors in the design of an immersive display," *IEEE Computer Graphics and Applications*, vol. 14, pp. 55–57, January 1994. 1.2

[8] A. Steed and C. Parker, "3d selection strategies for head tracked and non-head tracked operation of spatially immersive displays," *Immersive Projection Technology (IPT2004), Ames, Iowa*, 2004. 1.2

[9] A. Steed and C. Parker, "Evaluating effectiveness of interaction techniques across immersive virtual environmental systems," *Presence: Teleoper. Virtual Environ.*, vol. 14, pp. 511–527, Oct. 2005. 1.2

[10] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall, "Does the quality of the computer graphics matter when judging distances in visually immersive environments?.," *Presence: Teleoperators and Virtual Environments*, vol. 13, no. 5, pp. 560–571, 2004. 1.2

[11] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker, "Distance perception in real and virtual environments," *ACM Transactions on Applied Perception*, vol. 2, pp. 216–233, July 2005. 1.2

[12] M. H. Draper, H. A. Ruff, J. V. Fontejon, and S. Napier, "The effects of head-coupled control and a head-mounted display (hmd) on large-area search tasks," in *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, (Santa Monica, CA: Human Factors and Ergonomics Society), pp. 2139–2143, 2002. 1.2

[13] R. A. Ruddle, S. Payne, and D. Jones, "Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays?," *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 2, pp. 157–168, 1999. 1.2

[14] I. E. Sutherland, "A head-mounted three dimensional display," pp. 757–764, 1968. 1.2

[15] W. Robinett and J. P. Rolland, "A computational model for the stereoscopic optics of a head-mounted display," *Presence: Teleoperators and Virtual Environments*, vol. 1, no. 1, pp. 45–62, 1992. 1.2

[16] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson, "The effects of head-mounted display mechanics on distance judgments in virtual environments," *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pp. 35–38, 2004. 1.2

[17] T. Kawara, M. Ohmi, and T. Yoshizawa, "Effects on visual functions during tasks of object handling in virtual environment with a head mounted display," *Ergonomics*, vol. 39, no. 11, pp. 1370–1380, 1996. 1.2

[18] M. A. Goodale and A. Milner, "Separate visual pathways for perception and action," *Trends in Neurosciences*, vol. 15, no. 1, pp. 20–25, 1992. 1.3

[19] E. DeYoe and D. V. Essen, "Concurrent processing streams in monkey visual cortex," *Trends in Neurosciences*, vol. 11, no. 5, pp. 219–226, 1988. 1.3

[20] M. Goodale, C. Gonzalez, and G. Krliczak, "Action rules: Why the visual control of reaching and grasping is not always influenced by perceptual illusions," *Perception*, vol. 37, no. 3, pp. 355–366, 2008. 1.3

[21] R. Dyde and A. Milner, "Two illusions of perceived orientation: one fools all of the people some of the time, but the other fools all of the people all of the time," *Experimental Brain Research.*, vol. 144, pp. 518–527, 2002. 1.3

[22] L. E. Hughes, T. C. Bates, and A. Aimola Davies, "Grasping at sticks: pseudoneglect for perception but not action," *Experimental Brain Research*, vol. 157, pp. 397–402, 2004. 1.3

[23] G. Rizzolatti and M. Matelli, "Two different streams form the dorsal visual system: anatomy and functions," *Experimental Brain Research*, vol. 153, pp. 146–157, 2003. 10.1007/s00221-003-1588-0. 1.3

[24] F. Pavani, I. Boscagli, F. Benvenuti, M. Rabuffetti, and A. Farne, "Are perception and action affected differently by the titchener circles illusion?," *Experimental Brain Research*, vol. 127, pp. 95–101, 1999. 10.1007/s00221-005-0777. 1.3

[25] V. H. Franz, K. R. Gegenfurtner, H. H. Blthoff, and M. Fahle, "Grasping visual illusions: No evidence for a dissociation between perception and action," *Psychological Science*, vol. 11, pp. 20–25, 2000. 1.3

[26] A. Bayyari and M. E. Tudoreanu, "The impact of immersive virtual reality displays on the understanding of data visualization," in *Proceedings of the ACM symposium on Virtual reality software and technology*, VRST '06, (New York, NY, USA), pp. 368–371, ACM, 2006.

[27] A. Hogue, M. Robinson, M. R. Jenkin, and R. S. Allison, "A vision-based head tracking system for fully immersive displays," in *Proceedings of the workshop on Virtual environments 2003*, EGVE '03, (New York, NY, USA), pp. 179–187, ACM, 2003.

[28] K. Johnsen and B. Lok, "An evaluation of immersive displays for virtual human experiences," in *Virtual Reality Conference, 2008. VR '08. IEEE*, pp. 133–136, march 2008.

[29] J. Kjeldskov, "Interaction: Full and partial immersive virtual reality displays," in *Proceedings of IRIS24, University of Bergen, Bergen*, pp. 587–600, 2001.

[30] E. Lantz, "Future directions in visual display systems," *SIGGRAPH Comput. Graph.*, vol. 31, pp. 38–42, May 1997.

[31] E. Lantz, "The future of virtual reality: head mounted displays versus spatially immersive displays (panel)," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '96, (New York, NY, USA), pp. 485–486, ACM, 1996.

[32] L. McMillan and G. Bishop, "Head-tracked stereoscopic display using image warping," in *PROCEEDINGS SPIE, VOLUME 2409*, pp. 21–30, 1995.

[33] R. Messing and F. H. Durgin, "Distance perception and the visual horizon in head-mounted displays," *ACM Transactions on Applied Perception (TAP)*, vol. 2, pp. 275–280, 2005.

[34] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Trans. Information Systems*, vol. E77-D, no. 12, pp. 1321–1329, 1994.

[35] R. Raskar, "Immersive planar display using roughly aligned projectors," in *Proceedings of the IEEE Virtual Reality 2000 Conference*, VR '00, (Washington, DC, USA), p. 109, IEEE Computer Society, 2000.

[36] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The office of the future: a unified approach to image-based modeling and spatially immersive displays," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, (New York, NY, USA), pp. 179–188, ACM, 1998.

[37] G. E. Schneider, "Two visual systems.," *Science*, vol. 163, pp. 895–902, Feb 1969. 1.3

[38] N. Shibano, P. Hareesh, H. Hoshino, R. Kawamura, A. Yamamoto, M. Kashiwagi, and K. Sawada, "Cyberdome: Pc clustered hemi spherical immersive projection display," in

Proceedings of the 2003 International Conference on Artificial Reality and Telexistence, pp. 1–7, 2003.

[39] J. P. Wann, S. Rushton, and M. Mon-Williams, "Natural problems for stereoscopic depth perception in virtual environments," *Vision Research*, vol. 35, no. 19, pp. 2731–2736, 1995.

[40] F. Wanschitz, W. Birkfellner, M. Figl, S. Patruta, A. Wagner, F. Watzinger, K. Yerit, K. Schicho, R. Hanel, F. Kainberger, H. Imhof, H. Bergmann, and R. Ewers, "Computer-enhanced stereoscopic vision in a head-mounted display for oral implant surgery," *Clinical Oral Implants Research*, vol. 13, no. 6, pp. 610–616, 2002.

[41] M. Mon-Williams, J. P. Warm, and S. Rushton, "Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display," *Ophthalmic and Physiological Optics*, vol. 13, pp. 387–391, 1993.