8-17-2013

Near-Field Depth Perception in Optical See-Though Augmented Reality

Gurjot Singh

Near-field depth perception in optical see-though augmented reality

By

Gurjot Singh

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Computer Science
in the Department of Computer Science and Engineering

Mississippi State, Mississippi

August 2013

Near-field depth perception in optical see-though augmented reality

By

Gurjot Singh

Approved:

J. Edward Swan II
Professor of Computer Science and Engineering
(Major Professor)

T.J. Jankun-Kelly
Associate Professor of Computer Science and Engineering
(Committee Member)

Robert J. Moorhead
Professor of Electrical and Computer Engineering
(Committee Member)

Stephen R. Ellis
Committee Participant of Computer Science and Engineering
(Committee Member)

Edward B. Allen
Associate Professor of Computer Science and Engineering
(Graduate Coordinator)

Sarah A. Rajala
Dean of the James Worth Bagley College of Engineering

Name: Gurjot Singh

Date of Degree: August 17, 2013

Institution: Mississippi State University

Major Field: Computer Science

Major Professor: Dr. J. Edward Swan II

Title of Study: Near-field depth perception in optical see-though augmented reality

Pages of Study: 149

Candidate for Degree of Doctor of Philosophy

Augmented reality (AR) is a very promising display technology with many compelling industrial applications. However, before it can be used in actual settings, its fidelity needs to be investigated from a user-centric viewpoint. More specifically, how distance to the virtual objects is perceived in augmented reality is still an open question. To the best of our knowledge, there are only four previous studies that specifically studied distance perception in AR within reaching distances. Therefore, distance perception in augmented reality still remains a largely understudied phenomenon.

This document presents research in depth perception in augmented reality in the near visual field. The specific goal of this research is to empirically study various measurement techniques for depth perception, and to study various factors that affect depth perception in augmented reality, specifically, eye accommodation, brightness, and participant age.

This document discusses five experiments that have already been conducted. Experiment I aimed to determine if there are inherent difference between the perception of virtual and real objects by comparing depth judgments using two complementary distance judgment protocols: perceptual matching and blind reaching. This experiment found that real objects are perceived more accurately than virtual objects and matching is a relatively more accurate distance measure than reaching. Experiment II compared the two distance judgment protocols in the real world and augmented reality environments, with improved proprioceptive and visual feedback. This experiment found that reaching responses in the AR environment became more accurate with improved feedback. Experiment III studied the effect of different levels of accommodative demand (collimated, consistent, and midpoint) on distance judgments. This experiment found nearly accurate distance responses in the consistent and midpoint conditions, and a linear increase in error in the collimated condition. Experiment IV studied the effect of brightness of the target object on depth judgments. This experiment found that distance responses were shifted towards background for the dim AR target. Lastly, Experiment V studied the effect of participant age on depth judgments and found that older participants judged distance more accurately than younger participants.

Taken together, these five experiments will help us understand how depth perception operates in augmented reality.

Key words: augmented reality, mixed reality, perception, vision, proprioception

DEDICATION

To my parents Gurdeep Kaur & Gurwinder Singh, my godfather Gurcharan Singh, my family, as well as those without whom this would not have been possible.

ACKNOWLEDGEMENTS

Dr. J Edward Swan II

Dr. Stephen R. Ellis

Dr. T.J. Jankun-Kelly

Dr. Robert J. Moorhead II

Dr. J. Adam Jones

Shauncey Hill

Andreas Werner, ART Gmbh.

Department of Computer Science & Engineering

Institute for Neurocognitive Sciences & Technology

Center for Advanced Vehicular Systems

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

"A picture is worth a thousand words" (Brisbane [15])

The above quote aptly reflects the perceptual ability of human vision: very large and complex ideas can be conveyed with just one picture. Out of the five main senses, vision processes the incoming information at a much faster rate and in larger amounts. Therefore, it can be conclusively said that vision is the most important interface between humans and their surroundings.

The need for information exchange has never been critical as it is today. The advent of the internet and other media technologies has expanded the knowledge horizons of an individual by providing exposure to global events. "Every two days now we create as much information as we did from the dawn of civilization up until 2003" (Eric Schimdt [80]). This enormous amount of data requires sophisticated machines and methodologies to efficiently access specific information. The fields of computer visualization and human-computer interaction are driven by this need of efficient information flux between man and machine. The techniques and technologies in these domains strive to present data to a user in ways that are concise, complete, and intuitive.

A technology that has an enormous potential to act as on intuitive interface between man and machine is augmented reality. Augmented Reality (AR) is a display technology that enhances an individual's perception of their surroundings by adding additional information to their existing senses. In general, AR can refer to enhancement of any of the senses; however, for the purpose of this document, AR refers specifically to augmenting vision. AR augments vision by adding in-situ computer-generated imagery (virtual objects) in the view of the real world. The correct placement of a virtual object at its intended position requires knowledge of the location of the user as well as the location of the virtual object. Equipped with this information, a machine can present the virtual object with similar positional characteristics as a real object.

AR has numerous applications in a variety of areas, such as AR-assisted surgery, x-ray vision[1], manufacturing, maintenance, entertainment, and many more. The applicability of AR in these domains requires accurate presentation and perception of the virtual objects. However, the concept of accuracy is subjective. Different applications have different error tolerances and the required accuracy is relative to each functional context. For example, some accuracy requirements for surgical procedures are likely much more stringent than entertainment applications. To bring AR technology to a level where it can satisfy the requirements of a variety of applications requires answering the question: "how well can the virtual objects can be incorporated in an augmented environment?" This question encompasses two sub-questions; one each from the machine and user perspectives: First, how capable is current AR technology in creating and presenting virtual objects, as compared to the behavioral characteristics of a real object? And second, how well can a user perceive

[1]X-ray vision is ability to see through opaque objects by means of projecting or rendering the occluded infrastructure within or behind the occluding object.

the presented virtual objects and understand the intended spatial relationship among them and other real objects in the scene?

To date, most advancement in the AR domain has been made on the technical side, and it can be argued that AR technology has reached a level at which a virtual object can be rendered with a positional accuracy that is identical to a real object. However, even though technology has become advanced enough to solve the technical issues of AR, the human visual system is still imperfect in understanding key features of AR; such as the relationship between multiple layers of virtual objects and the relationship between virtual and real objects. So the question can be rephrased in part to ask how accurately can users perceive the depth and layout of the virtual objects in an augmented environment?

CHAPTER 2

BACKGROUND

2.1 Augmented Reality

Milgram’s Mixed Reality Continuum

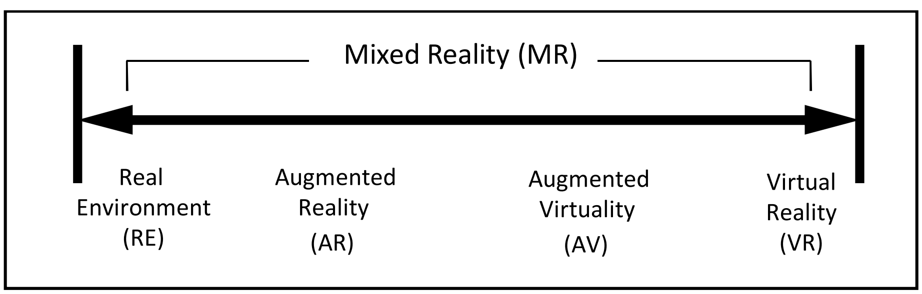


Figure 2.1

Mixed reality continuum (Milgram et al. [58]).

Based on various levels of augmentation, there are different levels of mixed reality such as *virtual reality* (VR), *augmented virtuality* (AV), *augmented reality* (AR), and the *real environment*, as categorized by Milgram et al. [58] as the mixed reality continuum (see Figure 2.1). In VR and AV, the environment around an observer is mostly virtual, consisting of none or a few real objects. In AR, the surrounding environment is primarily real with some virtual objects.

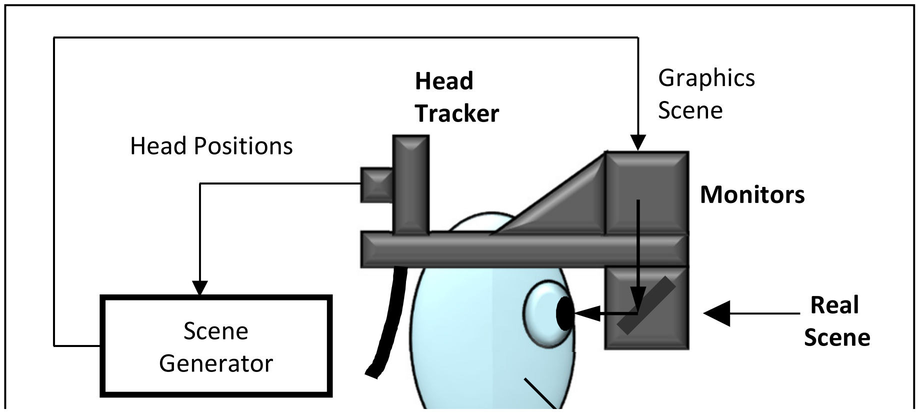
Azuma [5] gives a formal definition of an augmented reality system as one that satisfies the following three conditions:

- **Combines real and virtual**: The virtual information is seamlessly mixed with the real world.
- **Interactive in real time**: As the user moves, the virtual information updates in real time.
- **Registered in 3-D**: The virtual objects stay stationary with respect to the real world.

AR can be divided into two broad categories based on its implementation: *optical see-through* and *video see-through*. In optical see-through AR (see Figure 2.2), virtual objects are added to the view of an observer using partially transmissive optical combiners that retain a direct view of the real world while simultaneously showing the virtual objects superimposed on top of it. In video see-though AR (see Figure 2.3), virtual objects are added to a video feed of the real world scene captured using cameras, and the combined scene is presented either on a stereoscopic head-mounted display or on a monitor.

2.2 Depth Perception

Depth perception is an aspect of conscious experience when the human visual system (HVS) estimates the distance to an object. Depth perception is essential for the control of everyday spatially-related behaviors such as reaching, grasping, throwing, and intercepting



Video see-through AR

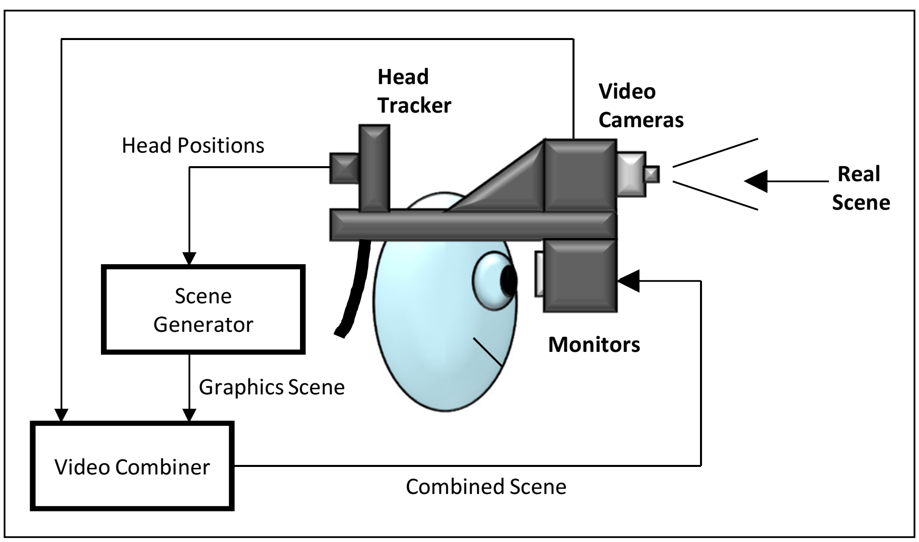


Figure 2.3

Video see-through AR system (Azuma [5]).

(Bingham [9], Mazyn et al. [54]). The seemingly simple process of reaching for an object with one's hand involves a number of complicated subprocesses. First, a perceptual mechanism uses vision to determine the distance to the object. After ascertaining the distance, the muscle control mechanism initiates a reach. Then, continuous feedback guides the hand to the object until the object is grasped. From this we can see that, when interacting with objects in the world, depth perception is the very first step. Depth perception is of two types:

- **Egocentric**: The distance of an object from an observer. The frame-of-reference that is used by the observer to ascertain distance lies at some point inside the observer.
- **Exocentric**: The distance between two objects as perceived by an observer. The frame-of-reference lies at some imagined point in space outside of the observer.

Even though the perception of depth is experienced only by the observer, the underlying perceptual mechanism that estimates egocentric distance is similar for all humans. First, the visual system captures a pair of two-dimensional retinal images of the scene, one from each eye. Afterwards, it identifies several visual stimulus variables in the scene, known as depth cues, which are then used to ascertain the distance relationship among various objects in the scene. As a result, a three-dimensional perceptual representation of the physical world is extracted from the flat and ambiguous retinal images.

2.2.1 Depth Cues

At least nine or more sources of depth information hve been identified, which allow an observer to understand the depth layout of the world (Cutting and Vishton [21], Cutting [20]).

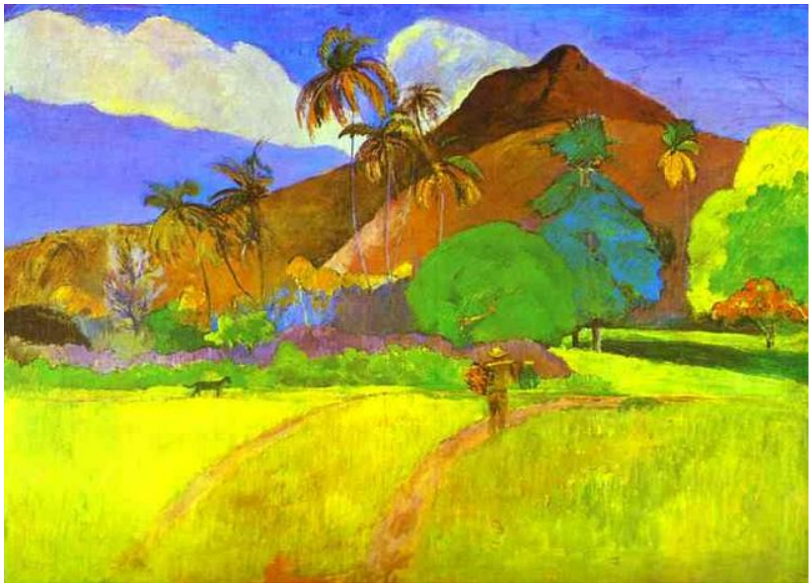


Figure 2.4

Occlusion: The closer trees occlude the view of the farther mountains (Paul Gauguin's Tahitian Landscape).

Occlusion (Figure 2.4) is a depth cue that is acquired from a scene when a closer object partially or completely occludes farther objects. It is a monocular depth cue, and can provide only ordinal information; it is not possible to measure any quantitative distance values on the basis of occlusion alone. However, it is a very strong ordinal depth cue and is effective at all distances without any attenuation; it works as well at far distances as it does at closer distances.

Height in the visual field (Figure 2.5) is the inverse relationship between an object's vertical position relative to the horizon as seen by the observer and its distance from the observer; objects closer to the horizon in the visual field are farther from the observer, and the objects farther from the horizon in the visual field are closer to the observer.

Figure 2.5

Height in the visual field: The closer objects are lower in the visual field (Jacob van Ruysdael's Haarlem).

Relative size (Figure 2.6) establishes depth relationships among objects of the same size. The more distant an object is, the smaller retinal extent it requires as compared to a similar-sized near object. By calculating the angular extent of similar objects it is possible to calculate distance ratio information. For example, if an observer sees two similarly-sized objects, with one object subtending half of the visual angle of the other object, the smaller object will be twice the distance of the larger one.

Relative density (Figure 2.6) compares the number of similar-sized objects or textures per solid visual angle. More objects will be seen per solid angle when they are at a farther distance than when they are closer. Therefore, the more objects per solid angle, the more distant they are. Similarly, a farther texture will appear more dense than a closer texture.

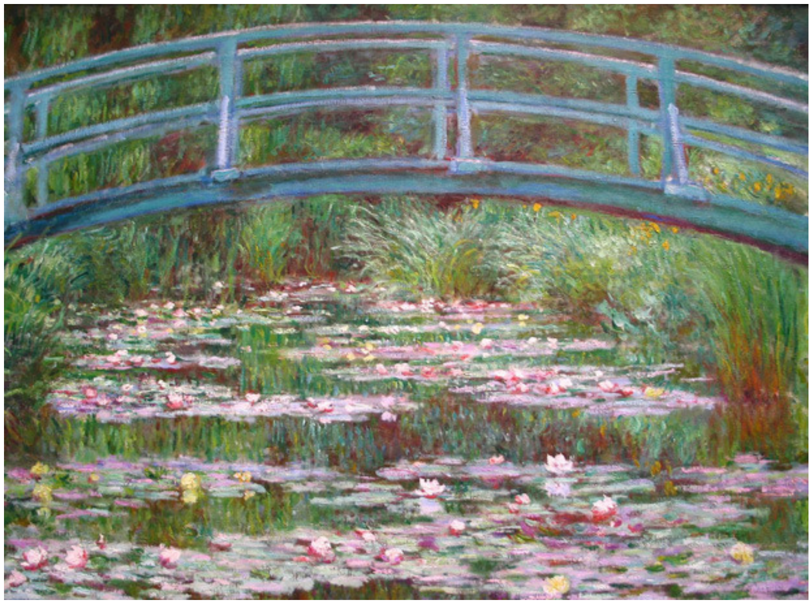


Figure 2.6

Relative Size and Density: The closer flowers look larger and sparser than the farther flowers (Claude Monet's Water-Lily Pond).

This cue can provide ratio information and it is possible to calculate the distance ratio between objects by counting the number of objects per solid visual angle. Essentially, relative size and relative density are two ways of understanding the same cue.

Aerial perspective (Figure 2.7) is the effect of atmospheric conditions on the appearance of objects. The atmosphere has suspended particles such as moisture, vapors, pollutants, dust, etc. Because of the presence of these particles, the contrast of an object decreases with distance and its color desaturates and shifts towards the atmospheric color. Based on this contrast calculations, this cue can provide quantitative depth information of the scene; however, it is effective only at large distances of 100 meters or more.

Figure 2.7

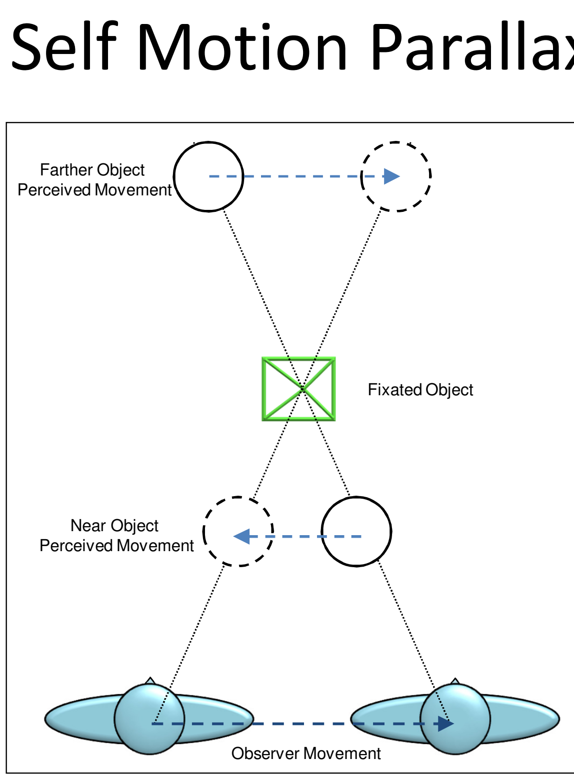


Figure 2.8

Motion Perspective: The change in the direction of apparent movement of objects with distance.

Motion perspective is the difference in the apparent motion of objects located at different distances from the observer. Motion perspective can be divided into two categories based on the movement of the observer. In the first category, the objects are fixed at certain distances and the observer is moving (see Figure 2.8). When an observer fixates on an object and moves sideways, the fixated object does not seem to move. However, objects closer than the fixated object appear to move quickly in the opposite direction of the observer, while farther objects appear to move slowly in the observer's direction of motion. In the second category, the observer is stationary while the objects are moving. For example, in a rain storm, the rain drops near the observer appear to fall more quickly than the drops farther from the observer.

Motion perspective is a special case of *motion parallax*, which is defined as the change in the rate of change of the line of sight with change in the position either of the observer or of the viewed object. The motion parallax is greater when the distance between the observer and the viewed object is smaller than when it is larger.

Accommodation (Figure 2.9) is the change in the shape of the crystalline lens in the eye to focus the light coming from an object onto the retina, and thus bring the viewed object in sharp focus. Objects which are in focus will look clear and sharp with well defined edges, while other objects which are closer or farther than the focused object will look blurred. The lens converges the light from the focused object exactly on the retina, while the light coming from the unfocused objects converges either in front of or behind the retina.

Convergence (Figure 2.10) is the rotation of the eyes to bring the visual axes to intersect at the viewed object. Our eyes are situated at different positions, and have different view

Convergence

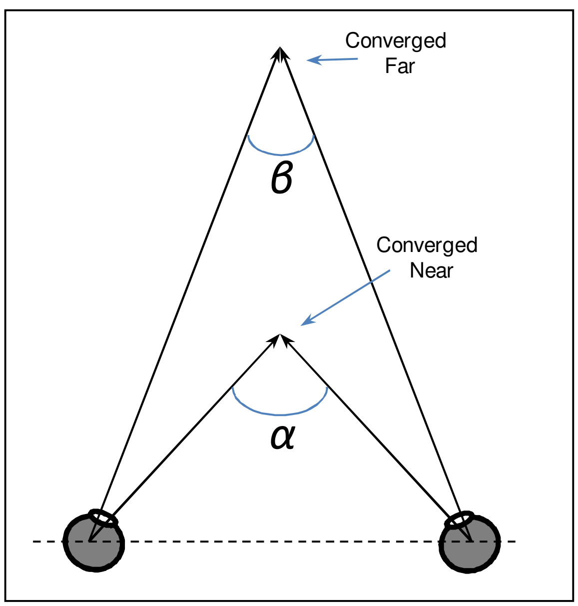


Figure 2.10

Convergence: The convergence angle is larger when converging at near objects than for far objects.

of the world from each other. To fixate on an object at a given depth, the eyes need to rotate inwards or outwards. Therefore, the angle between the lines of sight of two eyes is different for the fixations at different distances.

Figure 2.11

Binocular Disparity: Left and right eyes perspectives (Leonardo da Vinci's Mona Lisa).

Binocular disparity (Figure 2.11) is the difference between the relative position of an object on the retinal images in the two eyes (Wheatstone [97]). An object projects different images on each retina because the eyes are located at slightly different positions on the skull. This difference is called *disparity*. An object which is closer to the eyes will produce projections with more disparity than a more distant object. This disparity yields *stereopsis*, the impression of a solid space; this effect is very prominent for closer objects.

2.2.2 Depth Cues Categorizations

Depth cues have been categorized along different dimensions as shown in Table 2.1. Depth cues can be *binocular* or *monocular*; binocular cues use visual information from both eyes, while monocular cues require only one eye. Monocular cues are also known as pictorial cues, as they are used by artists to convey depth information in paintings and photographs.

Table 2.1

Depth Cues (Cutting and Vishton [21])

Depth cue	**Information**	**Ocularity**	**Type**	**Saliency (w.r.t. Distance)**
Occlusion	Ordinal	Monocular	Retinal	Constant
Height in Visual field	Quantitative	Monocular	Retinal	Decreases
Relative Size	Quantitative	Monocular	Retinal	Constant
Relative Density	Quantitative	Monocular	Retinal	Constant
Aerial Perspective	Quantitative	Monocular	Retinal	Increases
Accommodation	Ordinal	Monocular	Extra-retinal	Decreases
Convergence	Quantitative	Binocular	Extra-retinal	Decreases
Binocular Disparity	Quantitative	Binocular	Retinal	Decreases
Motion Perspective	Quantitative	Monocular	Retinal	Decreases

Depth cues are also categorized as *extra-retinal* or *retinal* based on their physiological linkage (see Table 2.1). Extra-retinal cues, such as eye vergence and accommodation, arise from human body sensors other than the retina, and are physiologically integrated in the human visual system. In principle, these cues can provide absolute distance information because of the physiological linkage. The visual system can interpret visual scenes binocularly by processing the relative changes of the projections of an object onto two retinal images with respect to the observer's stable inter-ocular distance. From this information, it is possible to triangulate the absolute position of the object. This absolute distance information is important to guide motor behavior while interacting with objects (Gillam [34]).

Retinal cues do not have any physiological linkage with the visual system. These cues rely solely on the visual information from the retinal images to provide distance information. Because there is no physiological linkage and therefore no internal scale to compare to, retinal cues require some additional information or some external reference to ascertain the overall scale of the visual sense. This additional information can be recalled from memory; e.g., the familiar size cue can provide distance information about previously seen objects. This information can also be provided by the existing objects in the environment itself, by triangulating the position of one object relative to the other objects in the scene. Retinal cues can only provide relative distance information of one object with respect to another object, and therefore covey no absolute distance information. However when combined with extra-retinal cues, retinal cues can provide definite distance information (Gogel and Tietz [36], Bingham and Pagano [11], Mon-Williams and Tresilian [61]).

Depth cues differ in their effectiveness based on the physical characteristics of the environment such as the scene content and the distance from the observer. Nagata [67] compared the effectiveness of various depth cues on a common scale called "depth sensitivity", which is the ratio of the viewing distance D to the detection threshold ΔD. They found that binocular disparity, accommodation, and convergence are mostly effective at closer distances, while relative size, brightness, and texture remain highly sensitive for more than 100 meters.

Later, Cutting and Vishton [21] compared different depth cues on a common scale of "saliency", which is the reciprocal of Nagata's [67] depth sensitivity. Depth saliency is the ratio of just noticeable distance ΔD to the viewing distance D. Figure 2.12 shows saliency of various cues relative to distance. Based on the saliency of different depth cues, Cutting and Vishton [21] divided the three-dimensional space around an observer into three different zones: *personal space*, *action space*, and *vista space* (see Figure 2.12). In near-field personal space, which extends up-to 2 meters, manipulation tasks such as reaching, touching, and grasping of objects can be performed. Here, convergence, accommodation, and binocular disparity are the most effective depth cues. In medium-field action space, which ranges from 2-30 meters, actions like throwing, navigation, and walking towards a target can be performed. Within this space, height in the visual-field and motion perspective are the most salient depth cues. Against both of these spaces, far-field vista space of more than 100 meters provides a general context and background. Here, the depth cue of aerial perspective is the most effective. Cutting also finds that the saliency of some cues such as occlusion, relative size, and density remain constant at all distances.

Later, Previc [75] categorized the space around an observer into four regions based on perceptual-motor operations. Apart from the three regions defined by Cutting and Vishton [21], he proposed one more region in between personal and action space called *focal-extrapersonal space*. This region is farther than personal space as the objects in this region can not be manipulated directly by reaching; however, the near-field cues such as convergence and accommodation can still be used to perform actions such as visual search and object identification.

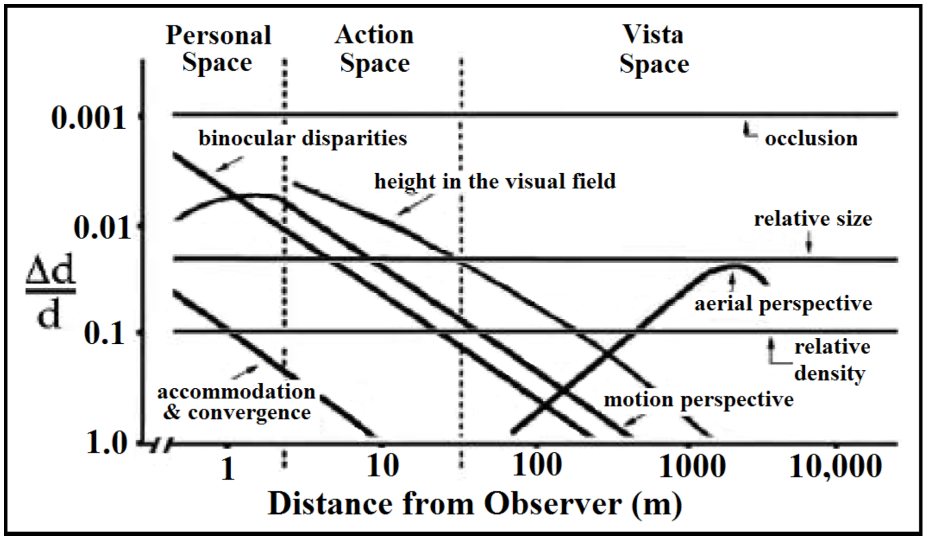


Figure 2.12

Saliency of various depth cues with distance.

2.2.3 The Near Triad

The work in this dissertation focuses on depth perception at near-field distances. When viewing an object in the near-field, the eyes converge (vergence), the lenses in the eyes become thicker (accommodation), and the pupils constrict (miosis). These three actions are interlinked physiologically and the mechanism of these three simultaneous reflexes is called the *near triad*. Classically, research on the near triad has followed two theories, favoring either accommodation or vergence as the primary stimulus or the driving factor for the other two reflexes (Maddox [52], Fincham and Walton [28]). Maddox [52] proposed that accommodation is the primary stimulus for vergence and pupil constriction. Later, Fincham and Walton [28] proposed the exact opposite, that vergence is the dominant stimulus for the near triad reflexes, and a single neuro-center controls both accommodation and vergence responses.

More recently, Semmlow and Hung [81] proposed a near triad model with separate accommodation and vergence control centers along with a cross-link interaction. This model suggests that near-field viewing involves accommodation and vergence affecting each other because of the cross-linking, and then the pupil constriction follows. Apart from the influence of accommodation and vergence, the changes in the pupil diameter also depend upon the scene illumination. These changes in pupil aperture should theoretically affect accommodation by changing the optical depth-of-field, however very little effect of change in pupil diameter on accommodation has been observed (Ripps et al. [76]).

In the near triad, convergence and accommodation are the main oculomotor cues and the link between them is known as the *convergence-accommodation reflex*. Because of this reflex, accommodation and convergence operate in unison; when attention changes from a distant object to a closer object, the eyes converge inwards to fuse the two retinal images of the fixated object, and simultaneously the lenses in the eyes change their shape to bring the fused image into focus. Furthermore, because of the cross-linking of the neural control centers for these reflexes, any changes in accommodation evoke changes in vergence (*accommodative vergence*) and any changes in vergence evoke changes in accommodation (*vergence accommodation*) (Fry [33], Fincham and Walton [28], Kersten and Legge [45]).

The reflexes in the near triad are affected by various factors. The rotation of the eyes (vergence) is stimulated by two factors: fusional vergence and accommodative vergence.

onvergence-Accommodatio

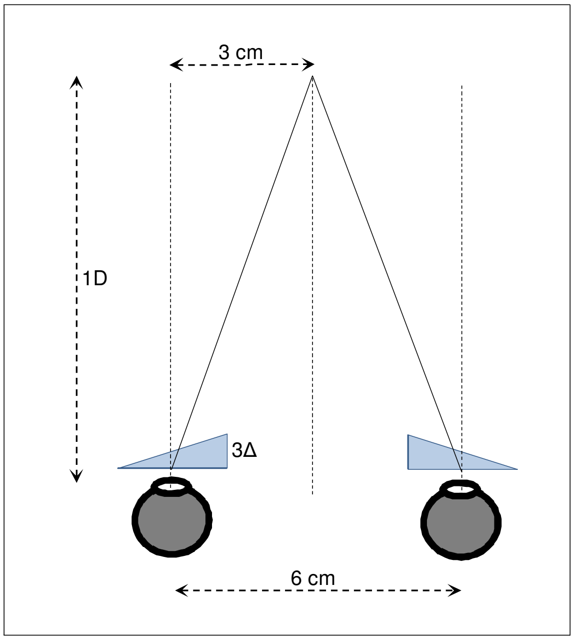


Figure 2.13

Convergence-accommodation relationship.

The accommodative demand or the focusing of the eyes is typically measured in diopters (D), which is the reciprocal of the distance of the target object from observer's eyes in meters ($D = 1/m$). Therefore, a target at 0.5 meters would yield an accommodative demand of $2D$. The convergence demand or the rotation of eyes is measured in prism diopters (Δ), and it is determined by the target distance as well as the inter-pupillary distance (IPD) of the observer. One prism diopter is defined as the deviation of 1 cm at the distance of 1 m. Therefore, as shown in Figure 2.13, an observer with an IPD of 6 cm would yield a total convergence demand of 6Δ (3 cm deviation, or 3Δ for each eye), while converging on a target at 1 m in front of their eyes. Due to the convergence accommodation reflex, the convergence demand and the accommodative demand are connected and can be calculated as (Peli [73]):

$$\textit{Convergence Demand } (\Delta) = \textit{Accommodation Demand (D)} \times \textit{IPD (cm)} \tag{2.1}$$

The convergence demand can also be measured in degrees as:

$$\textit{Convergence Demand (degrees)} = tan^{-1}\left(\frac{\Delta}{100}\right) \tag{2.2}$$

Another unit that measures ocular convergence is the *meter-angle*, which is defined as the amount of convergence required to binocularly view an object at 1 m while exerting 1 D of accommodation. As shown in Figure 2.14, the value of a meter-angle depends upon one's inter-pupillary distance (IPD), and therefore is unique for every individual. The meter-angle is defined as:

$$\textit{1 meter-angle } (\theta) = 2 * tan^{-1}\left(\frac{IPD}{2}\right) \tag{2.3}$$

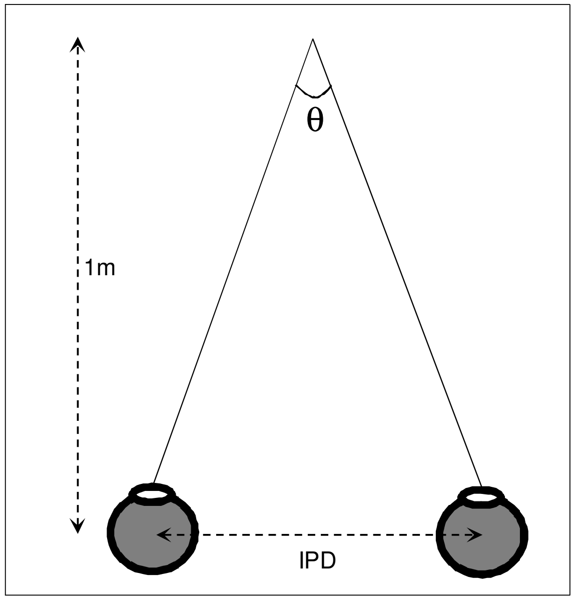


Figure 2.14

Meter-angle (θ).

2.2.4 Depth Cues Integration

The process of distance estimation involves combining the distance information from all of the available cues to define a single percept of distance. Anderson [3] describes distance perception as a three-step process: (1) a *psychophysical transform*, which transforms the information from different depth cues into their respective distance signals, (2) a *psychological integration process*, which combines all the distance signals into a final distance signal, and (3) a *psychomotor transform*, which converts the final distance signal into a response-indicating distance. Mathematically, this definition of depth perception can be explained by the theory of information integration (Anderson [2], Bruno and Cutting [16]), where information from various depth cues $c_1, c_2, ..., c_n$ can be combined in an integration model described as:

$$d = f\ (\ c\ [w_1 * c_1, w_2 * c_2, ..., w_n * c_n]) \tag{2.4}$$

where d is the distance to be determined, $w_1, w_2, ..., w_n$ are the weights assigned to each depth cue, c is the combination or integration rule, and f maps the combined information to a response.

Even though many depth cues have conclusively been identified, it has not been fully understood how different depth cues are integrated by the human visual system to provide a single estimate of egocentric distance. Previous empirical depth perception studies suggest that during distance estimation, adding additional depth cues generally increases the amount of depth seen and the consistency and accuracy of the distance judgments (Künnapas [48]).

Many depth cue integration models have been proposed. Bruno and Cutting [16] proposed that information from various depth cues is gathered by separate visual minimodules and processing systems, and then this data is combined together in the simplest manner of addition; one cue can substitute for another, and the more cues are present the more depth is revealed. Massaro [53] proposed the fuzzy logic model of perception (FLMP) in contrast to the integration model proposed by Bruno and Cutting [16]. This model suggests that the cue integration process is non-additive and a cue with more certainty makes a bigger contribution as compared to the other ambiguous cues.

Landy et al. [49] proposed three different models for depth cue integration. The first one is the *weak observer model*, where depth cues are considered to provide independent depth maps about various objects in a scene. The information from different depth cues is combined by averaging the depth map values of all of the cues resulting in an overall depth map of the scene. Even though this model is simple and modular, it is susceptible to problems such as the lack of normalization among the depth maps, saliency, reliability, and the availability of different depth cues. The second model is the *strong observer model*, which promotes interactive and holistic processing of depth cues. It assumes that the most-probable three-dimensional interpretation of a scene depends on the current retinal data. In this model, what constitutes a distinct depth cue is not given in advance and it evolves gradually and is tested as a part of the depth cue combination model. Various depth cues "mysteriously" interact with each other and the combination rules are not necessarily linear. One important issue with the strong observer model is that it cannot be falsified by an experiment and is therefore not experimentally testable. The third model, and what Landy et al. [49] prefer, is the *modified weak fusion model*, which uses the modular view of depth calculations (a weak observer property) but puts restrictions on the interaction (a strong observer property) by defining how various depth cues provide depth maps and how various depth maps are combined. Every depth cue provides qualitatively different information and this model combines information from multiple cues using weighted averaging. At its core, the modified weak fusion model is a weighted averaging of multiple depth cues based on a cue's reliability as defined by robustness of estimate, availability of cue, and information provided by ancillary measures.

2.3 Depth Cues in Augmented and Virtual Reality

At near-field distances, three depth cues that play major role in depth perception are: binocular disparity, vergence (convergence), and accommodation. In AR/VR displays,

the

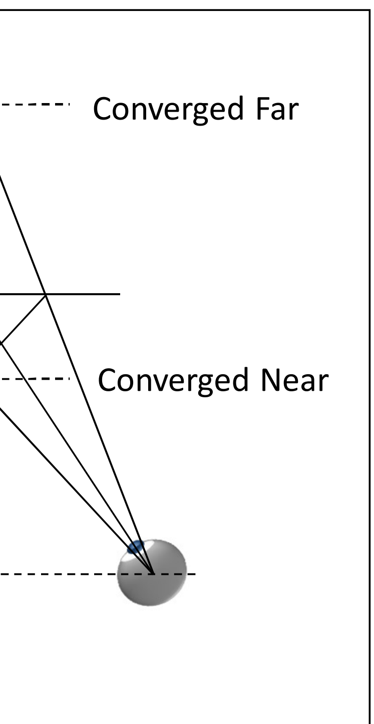
es in

ence-Accommodation Mismatch

opic

and

the



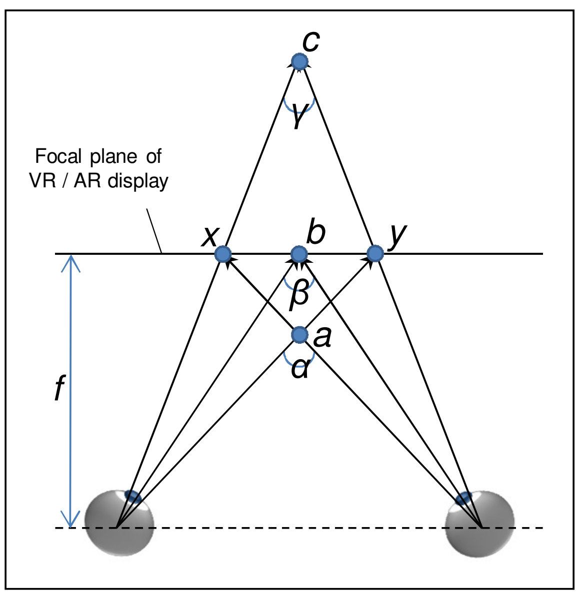


Figure 2.15

Convergence-accommodation mismatch.

In real world, accommodation and convergence operate in unison and depend upon viewing distance. Normally, when one switches attention across different distances, because of the convergence-accommodation reflex, convergence and optical power of the eyes changes to maintain binocular vision and a clear sharp image of the scene (Owens and Leibowitz [69], Mon-williams and Tresilian [62]). For farther objects, the eyes diverge outwards and the lens in the eye thins to avoid double vision and a blurry image. As shown in Figure 2.15, when the eyes switch attention from b to c, the vergence angle changes from β to γ, and the accommodation changes from b to c. For closer objects, the eyes converge inwards and the lens thickens to maintain a single sharp image of the object; when the eyes switch attention from b to a, the vergence angle changes from β to α, and the accommodation changes from b to a.

Normally, these changes in eye convergence and accommodation happen under unconscious control. The human visual system developed the convergence-accommodation coupling to view objects with correct vergence and accommodation demands. This coupling remains in effect even without an external stimuli; when one eye is closed and the other eye accommodates at a closer object, then under the effect of this coupling, the closed eye rotates to converge at a closer distance (Muller [65], Peli [73]). Then, if the closed eye is opened, only a small amount of correction is needed to fuse the retinal images, in addition to the already present accommodative-vergence. Therefore, when the eyes accommodate, they converge even without any external convergence stimuli (Gillam [34]).

In stereoscopic displays, the normal coupling between accommodation and convergence is disrupted (Wann et al. [96]). As shown in Figure 2.15, even though the virtual objects are at three different depth (a, b, and c), they are rendered on the screen at the distance f from the eyes. Even though the convergence changes when looking at virtual objects at different depths, the accommodation stay fixed at the distance f. Therefore, a fixed-accommodation display forces the visual system to always focus at one distance even if the eyes are converging at a different distance. This need of fixed accommodation while changing eye vergence disrupts the normal accommodation-vergence coupling. This problem is known as *convergence-accommodation mismatch* (Drascic and Milgram [22]), and is known to cause depth misjudgments, fatigue, and visual stress in virtual environments (Bingham et al. [10], Hoffman et al. [40]).

2.4 Depth Judgment Techniques

The human visual system perceives the external physical world by capturing two retinal images; one from each eye. From these retinal images, the visual system extracts a three-dimensional internal representation of the world called *visual space*. Therefore, visual space is that component of the physical world that is acquired by visual perception (Loomis and Knapp[51]). The term "depth perception" usually refers to ascertaining depth relationships among objects in perceived visual space.

The measurement of depth perception is a problem, as it is an invisible cognitive state that is experienced only by the observer, and therefore cannot be measured directly by an experimenter. To measure depth perception, another component of the physical world called *action space* is defined, which is a component of the physical world in which actions can be performed that result in some measurable behavior. Psychologists assume that visual space is the primary determinant of any spatial behavior in action space, and that actions performed in action space will represent the properties of visual space. When a distance judgment task is performed in action space that results in some measurable response, it can quantitatively measure distance perception in visual space.

Scientists have used two types of techniques to measure distance perception: *perception* techniques and *perception-action* techniques (Bingham et al. [10]). Both of these techniques involve visually perceiving an object and estimating its egocentric distance; however, the way the perceived distance is measured is different. Perception-based techniques require active cognition and involve object identification, recognition, and the ability to verbally describe the distance. An example of a perception-based technique is *verbal report*. In verbal report, the distance to a stimulus object is visually estimated, and then it is conveyed verbally in some familiar distance units (feets, meters, arm-lengths etc.). Perception-action techniques involve motor actions to indicate the perceived distance. An example of a perception-action technique is *visually directed motor behavior*, where an observer performs a physical action to indicate the distance to a visually perceived stimulus (Knapp [46]).

Like any control system, the techniques that are used to measure depth perception are susceptible to errors. Depth perception errors consists of perceptual errors and motor program errors (Woodworth [102]). Perceptual errors involve incorrect depth estimation, and the scope of these errors is in visual space. Motor program errors result from incorrect movement instructions to effectors (in context of our experiments, effectors refer to muscles and joints). These errors affect task performance in action space.

Out of perception-based and perception-action based tasks, perception-based tasks are less accurate. For example, the perception-based task of verbal report has been found to be confounded by cognition with high response variability (Foley [30], Gogel [35], Mon-Williams and Tresilian [60], Swan et al. [84]). Previous studies that compared verbal report and visually directed reaching found no correlation between the two distance measures, even though the two tasks were performed within the same trial (Pagano and Bingham [71], Napieralski et al. [68]).

Compared to verbal report, perception-action tasks are more accurate. The difference between the accuracy of these measures can be attributed to anatomical differences of the control centers of the two tasks. Goodale and Milner [38] found that perception-based judgments (object identification, recognition, and verbal description) and perception-action based judgments (reaching and grasping) are physiologically dissociated and are mediated by distinct neurological channels (A more recent survey about the relationship between visual perception and action can be found in Goodale [37]). Bridgeman et al. [14] found that the perception-action channel is more accurate at estimating distance, whereas the perception channel is more prone to distorted perceptual judgments. Bingham et al. [10, 12] proposed that the high accuracy of visually directed motor tasks could be a result of the regular calibration of the visio-motor system resulting from feedback from everyday experiences of manipulating objects in the real world.

Perception-action tasks can further be divided into two categories based on available feedback: *closed-loop* tasks and *open-loop* tasks. The terms "closed-loop" and "open-loop" are from the field of control theory, and they represent the behavior of a dynamic

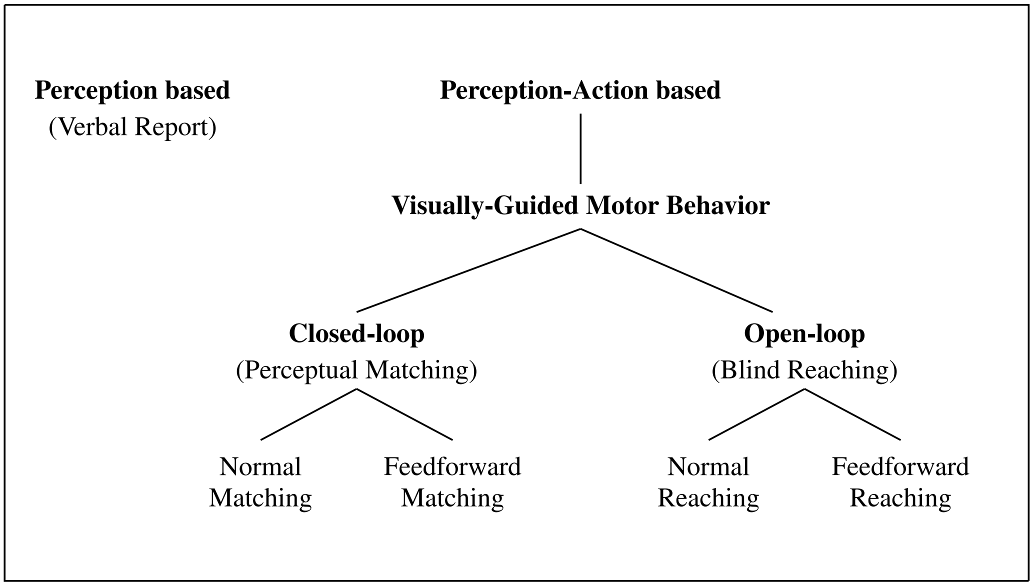


Figure 2.3

Figure 2.16

Depth judgment techniques.

Depth judgment techniques

system under the effect of feedback. In a closed-loop system, input to a system is adjusted based on a feedback signal. Therefore, the output is defined by the input signal as well as the feedback. In an open-loop system, there is no feedback and the output is defined solely by the input signal (Levine [50]). For our experiments, we used a closed-loop task of *perceptual matching* and an open-loop task of *blind-reaching* (see Figure 2.16).

18

2.4.1 Perceptual Maching

The perceptual matching task is based on our everyday experiences of real world interactions. In the real world, people interact with objects using closed-loop perception; for example, keeping one's eyes on a cup while reaching for it. They execute the reaching action under the direct perceptual guidance of the feedback information obtained from looking at the cup and their hand. Because of the natural perceptual-motor coupling in a closed-loop task, the visual feedback serves as a perceptual basis for the matching action.

A perceptual matching task involves matching the distance to a target object by pointing at it with a pointer object. An observer moves a pointer towards the target object until both of the objects are at the same egocentric distance from the observer. Depending upon the experimental context, the pointer object is moved either directly by the observer using some mechanical equipment (Ellis and Menges [26], Singh et al. [83]) or indirectly by an experimenter (Swan et al. [85], Wu et al. [103]). Depending upon the spatial constraints, the pointer object can be a physical object or the hand of the observer.

A perceptual matching task involves two subprocesses: visual perception and a matching action. The visual perception process involves visually eliciting the information about the difference between the perceived locations of the pointer and the target object. The matching action involves moving the pointer to reduce the difference in the perceived depth of the pointer and the target object. The visual feedback from the objects directs the movement of the pointer, and this movement in turn affects the current state of the objects. Therefore, the perception of the objects and the movement action of the pointer affect each other in a closed loop. Along with visual feedback, perceptual matching also involves proprioceptive feedback, which is the sense of limb positions. Proprioceptive feedback is available when an observer performs the perceptual matching task using their hand as the pointer object. Even when a physical pointer is used, some amount of proprioceptive feedback is still available.

There are two variations of the perceptual matching task: *normal matching* and *feedforward matching* (see Figure 2.16). In a normal matching task, the target object as well as the pointer object are visible to the observer at all times. In feedforward matching, the target object is visible only at the start of the trial and is hidden during the matching task. The former task has little or no memory requirements, while the later task requires the observer to remember the perceived position of the target object during the matching task.

Perceptual matching techniques have been widely used for real world depth judgments (Prablanc et al. [74], Bingham et al. [10]), as well as in augmented reality environments (Ellis and Menges [26], MacCandless et al. [55, 56], Rolland et al. [77, 78], Singh et al. [83]). In these studies, a real object was used as a pointer object to match either a real object (in the real world) or a virtual object (in AR). It is important to note that this definition of perceptual matching does not hold in virtual reality (VR), because VR can not provide the direct view of a real pointer.

Even though the perceptual matching tasks are more suitable for real world applications, some scientists do not consider them as appropriate measure of definite depth perception (Bingham and Pagano [11], Prablanc et al. [74]). The visual feedback available during the matching task can render this task more of a disparity matching task, rather than a perceptual matching task, where one object is matched relative to another object. Therefore, this technique gives a relative measure of depth rather than an absolute measure of egocentric depth (Bingham and Pagano [11]). It is not necessary to have an internal representation of the depth or the absolute distance information to the target object to perform this task. Prablanc et al. [74] explains the perceptual matching task as a combination of three different types of controls that operate during optimal visual guidance of the arm towards a target: (1) a central programming triggered by the retinal input (visual image of the target), (2) a central processing of peripheral information to program initial hand-motor coordination, and (3) a peripheral processing of the retinal information of the position of the hand and the target during the movement. This explanation is in line with Bingham et al. [11] about the perceptual matching task being more of a retinal disparity task rather than a representation of the consciously experienced visual space.

Nonetheless, this task is important because of its real world applications and the fact that it can serve as a ground truth or best possible performance for depth judgments, even if the results only measure relative depth perception. For example, a perceptual matching task is most appropriate in an image-guided surgery that includes using a scalpel for cutting to the depth of an AR presented marker inside a patient's body. Its suitability to act as ground truth is supported by the fact that it uses both visual and proprioception feedback, which when combined are found to give better depth judgment performance than if either of them is used alone (Van Beers et al. [90]). Also, real-world perceptual matching is extremely accurate at reaching distances.

2.4.2 Blind Reaching

As discussed in the previous section, many do not consider perceptual matching to be an appropriate measure of absolute distance perception. Another category of perception-action tasks are open-loop, which are considered as a more valid measures of absolute distance perception. These tasks involve judging the distance to a target object with a hidden pointer without any corrective mechanism of feedback.

At near field distances, an open-loop task to measure depth perception is *blind reaching*, where an observer reaches to a target object to indicate its distance with a hidden pointer. Because there is no visual feedback available, observer has to rely on some internal sense of the perceived distance to perform this task. In blind-reaching, even though there is no visual feedback present, the proprioceptive feedback is still available, and therefore this task is more precisely a visually open-loop task.

Blind-reaching tasks are of two types: *normal reaching* and *feedforward reaching* (Bingham et al. [10]) (see Figure 2.16). In normal reaching, the target object is visible to the observer both before and during the reach. In feedforward reaching, the target object is viewed before the reach begins, and then it is hidden from the view of the observer. Therefore, the reach is executed based on the remembered location of the target.

Blind-reaching technique has been widely used in the depth perception studies at near-field distances for both real and virtual environments (e.g., Bingham et al. [12]; Mon-Williams and Bingham [59]; Mon-Williams and Tresilian [61, 62]; Tresilian et al. [88]; Naceri et al. [66]; Napieralski et al. [68]). While some studies have found very accurate responses for reaching (Mon-Williams et al. [64], Mon-Williams and Tresilian [61], van Beers et al. [90], von Hofsten and Rosblad [92], Wann [94]), other studies have found far less accurate responses with median errors of up to 25 cm (Foley and Held [32], Foley [30])

Blind reaching does not involve visual feedback, and therefore the reaching judgments represent some internal sense of depth. However, blind-reaching does utilizes proprioceptive feedback, which has been found to be pliable and easily susceptible to drift in the absence of some explicit correction mechanism. Harris [39] studied adaptation to prism-produced displacement of the visual field and found that when visual and proprioceptive senses provide discrepant information, it is the proprioceptive sense that changes. Wann et al. [95] studied depth perception using a blind-reaching and found that the participants had a tendency to underestimate the distance by bringing their judging finger closer to their body. Paillard and Brouchon [72] also found that when using only proprioception, the perceived location of an unseen hand drifts as quickly as in 12 seconds. These findings suggest that in the absence of some corrective mechanism the proprioceptive sense drifts and the response variability increases.

Because the proprioception is susceptible to drift over time, Bingham and Pagano [11] advocate using perception-action tasks with feedback while evaluating definite depth perception. The availability of feedback provides a standard of accuracy, where one tends to achieve an optimized response by minimizing the response errors over repeated trials. The removal of the feedback renders the system unstable and results in decreased accuracy over time. Mon-Williams and Bingham [59] studied the effect of haptic feedback on the perceived position of a virtual object by providing perturbed haptic feedback. They found that the proprioceptive sense is indeed very flexible, and calibration resulting from haptic feedback at one position transfers to the responses made at other positions in a proprioception-only task. This intentional feedback can be used advantageously to correct any initial inconsistencies between the actual position and the perceived position of the viewed objects.

2.5 Light Measurement

The human visual system perceives the external world by capturing the visible light portion of the electromagnetic spectrum. The electromagnetic spectrum extends from radio waves (with wavelength $>$1 m) down to gamma rays (with wavelengths of $<$ 0.01nm). Visible light covers only a small portion of this spectrum (with wavelengths of 380$-$770 nm), and exhibit ray, wave, and quantum properties.

The field of *radiometry* deals with the measurement of the energy of electromagnetic waves, over the entire electromagnetic spectrum. This energy is called *radiant energy* (Q_e), and is measured in *joules* (J). The total amount of radiant energy produced by an object per unit time is known as *radiant flux* or *radiant power* (ϕ_e), and is measured in *watts* (W). Assuming a point source, the radiant power is quantified over solid angles (steradians). This measure is called *radiant intensity* (l_e), and is measured in *watt per steradian* ($W \cdot sr^{-1}$). Another measure that quantifies the electromagnetic energy over a source area is *radiance* (L_e), and is defined as radiant intensity per unit projected source area ($l_e \cdot m^{-2}$ or $W \cdot sr^{-1} \cdot m^{-2}$).

A subset of radiometry is *photometry*, which deals with the measurement of visible light as perceived by the human eye. The energy of the light waves in the visible spectrum is called *luminance energy* (Q_v), and is measured in *lumen seconds* ($lm \cdot s$). The total amount of luminance energy produced by an object per unit time is known as *luminance* *flux* or *luminance power* (ϕ_v), and is measured in *lumens* (lm). Therefore, luminous power is the total perceived power emitted in all directions. Another unit that measures the luminance flux in a particular direction is called *luminance intensity* (l_v), and is defined as the perceived power per unit solid angle (steradian). Luminance intensity is measured in *candela*, and is defined as lumen per steradian ($cd = lm \cdot sr^{-1}$). The luminous intensity per unit area of light traveling in a given direction is known as *luminance* (L_v), and is measured in *candela per unit area* ($cd \cdot m^{-2}$)

Table 2.2 shows the corresponding radiometric and photometric measures.

Table 2.2

Radiometric and Photometric measures.

Radiometric	Photometric
Radiant Energy (Q_e) joule (J)	***Luminous Energy*** (Q_v) lumen second ($lm.s$)
Radiant Flux or Radiant Power (ϕ_e) watt (W)	***Luminous Flux or Luminous Power*** (ϕ_v) lumen (lm)
Radiant Intensity (l_e) watt per steradian ($W \cdot st^{-1}$)	***Luminous Intensity*** (l_v) lumen per steradian ($lm \cdot st^{-1}$) candela (cd)
Radiance (L_e) radiant intensity per unit area ($l_e \cdot m^{-2}$) watt per steradian per unit area ($W \cdot st^{-1} \cdot m^{-2}$)	***Luminance*** (L_v) luminous intensity per unit area ($l_v \cdot m^{-2}$) candela per unit area ($cd \cdot m^{-2}$)

2.5.1 Radiometric to Photometric Conversion

The human eye's sensitivity varies over the different wavelengths in the visible spectrum. Photometry takes this sensitivity change into account by weights that represent the sensitivity of the human eye for different wavelengths. In 1924, the CIE (Commission Internationale de l'Eclairage) defined a photopic luminosity function that provides a standardized model of the eye's response to light as a function of wavelength. The measurement quantities in photometry are wavelength-weighted based on this model.

The relationship between the photometric measure of luminance and the radiometric measure of radiance is defined as (Ryer [79]):

$$L = 683 * \int_{\lambda=360nm}^{\lambda=830nm} R(\lambda) P(\lambda) \mathrm{d}\lambda. \tag{2.5}$$

Here L is luminance of the target object measured in *candela per sq. meter* ($cd \cdot m^{-2}$), $R(\lambda)$ is the radiance energy measured in *watts per steradian per sq. meter* ($W \cdot sr^{-1} \cdot m^{-2}$), P is the photopic luminous efficiency function, and λ is the wavelength of visible light measured in nanometers.

CHAPTER 3

PREVIOUS WORK

3.1 Depth Perception in the Real World

Depth perception has been an important area of psychological research for over 100 years. A large amount of literature is available on how depth perception works in the real environment. The overall focus of depth perception studies has been on three main fronts: identifying various depth cues in a scene and their effectiveness based on scene properties, depth cue integration, and psychophysical measurement of depth perception.

The focus of this dissertation is on depth perception at near-field distances. Within this range of distances the main depth cues are eye vergence, accommodation, and binocular disparity (Cutting [20]).

3.1.1 The Role of vergence in depth perception

At near-field distances, vergence has been conclusively found to provide egocentric depth information (Fincham and Walton [28], Foley and Held [32], Owens and Liebowitz [70], Brenner and Van Damme [13], Viguier et al. [91]). However, the contribution of vergence to depth estimation depends upon the scene properties. In reduced-cue-conditions, vergence is an important depth cue for depth perception (see Foley [31] for a comprehensive review). However, its role in depth perception becomes complex when other retinal cues are available. In low cue conditions, vergence-specified distance is relative to a resting vergence (Owens and Liebowitz [70]) i.e., participants tend to judge distance towards their resting vergence; an object closer than the participants' resting vergence is overestimated and an object farther than the participants' resting vergence is underestimated.

Depth perception is also affected by the remembered position of a previously judged target (Brenner and Van Damme [13]). If the viewed target is closer than the previously judged target, the participants tend to overestimate its distance, and if the viewed target is farther than the previously judged target, the participants tend to underestimate its distance. Vergence has been found to be most effective up to a distance of 2 meters; as the distance to the target increases, vergence becomes unreliable for judging distance (Tresilian et al. [88]). When other depth cues are available in the scene, the amount of contribution that vergence makes towards depth estimation reduces significantly. Turvey and Solomon [89] found that even at near-field distances, vergence does not contribute to distance perception when other retinal information is sufficient for determining distance.

3.1.2 The Role of accommodation in depth perception

The role of accommodation as a source of depth information has been thoroughly studied. Classically, accommodation itself was not considered a strong source of depth information (Künnapas [48], Kenyon et al. [44], Mon-Williams et al. [63]). However, it has been found that accommodation influences depth perception indirectly by affecting vergence through the convergence-accommodation reflex (Semmlow and Hung [81], Mon-Williams and Tresilian [62]).

Even though the role of accommodation as a sole source of depth information remains inconclusive, some studies suggest that accommodation in itself can act as a source of depth information, at least for some observers (Wallach and Floor [93], Fisher and Ciuffreda [29], Mon-Williams and Tresilian [61]). Fisher and Ciuffreda [29] found a correlation between accommodation and distance responses for some individuals. Mon-Williams and Tresilian [61] also found similar results; the pointing responses of two out of six participants showed a strong relationship of accommodation to target distance, two showed a weaker relationship, and two did not show any relationship. Together, these studies suggest that accommodation affects depth perception indirectly by influencing vergence.

3.1.3 Age and depth perception

As a person ages, the accommodation ability of their eyes diminishes. This condition is called presbyopia (Duane [23], Beers and Heijde [6], Kasthurirangan and Glasser [43]). Because of this condition, people typically need reading glasses at some point in their 40s. A number of factors has been found to cause presbyopia, such as hardening of the crystalline lens, increase in the lens thickness and diameter with age, changes in the angle of the zonular attachments to the lens, age-related decrease in ciliary body movement, and reduced elasticity of the choroid (Kasthurirangan and Glasser [43]). Presbyopia normally starts at the age of 20, and by age 60 the lens is completely incapable of accommodating to an object at near-field distances (Duane [23], Bell et al. [7], Whitbourne [98]). Since accommodation is a source of distance information, any changes in normal functioning of accommodation is bound to change the way distance is estimated by older people.

Bian and Anderson [8] studied the effect of age on the perception of egocentric distance for medium-field distances of 4-12 meters. They found that the older participants judged the egocentric distance more accurately as compared to the younger participants, who underestimated and showed distance foreshortening. They also found that this age-related difference was not due to the difference in distance scaling, difference in output calibration, or use of eye-height or texture gradient information. They suggested that the accurate distance perception in older participant might have resulted from their experience over life-time, difference in the perceived slant of the ground surface, or a greater reliance on pictorial cues with increased age.

3.1.4 Brightness and depth perception

Brightness has been used to convey depth in art and paintings for a long time. Leonardo Da Vinci explains the relationship between brightness and depth as:

> *"Among bodies equal in size and distance, that which shines the more brightly seems to the eye nearer and larger"– Leonardo Da Vinci* (McCurdy [57])

The effect of brightness on depth perception has been studied thoroughly in the real environments (Ashley [4], Coules [18], Farne [27], Egusa [24, 25]). In all of these studies, the general finding was that among objects situated at the same fronto-parallel plane, the brighter object appears closer to the observer. This effect was present for both monocular and binocular conditions, as well as for both near-field (Ashley [4], Farne [27]) and medium field (Coules [18]) distances.

Ashley [4] suggested that a brighter object stimulates a larger tract of retina, resulting in an increase in the perceived size of the object, and thus the object appears closer because of the relative size cue. Another explanation of this phenomenon is based on the near-triad reflexes, which causes contraction of the pupil while looking at near objects due to changes in convergence and accommodation. Since the increased brightness also causes the pupil to contract, a brighter object is perceived to be closer. Tayler and Sumner [86] suggested that brighter objects stimulate the retina more intensely, and hence give more vivid images. Usually closer objects are more vivid, and therefore the brighter objects are perceived to be closer.

Brightness contrast of an object with respect to the background also affects depth perception. If there is a difference between brightness contrasts, the object with the largest brightness contrast is perceived to be the nearest; when the background is darker the brighter object is perceived as the nearest, and when the background is brighter the darker object is perceived to be the nearest (Egusa [24], Farne [27]). Egusa [24] studied the effect of object presentation on depth perception by presenting a target either as a foreground or as a background. They found that the target with the greatest brightness contrast was judged closer when it appeared as a foreground image, and it was judged farther when it appeared as a background.

3.2 Depth Perception in Virtual Reality

Over the past 15 years, more than 40 studies have examined egocentric depth perception in virtual reality environments at medium-field distances (Loomis and Knapp [51], Swan et al. [84]). Both of these references contain extensive literature surveys about depth perception in virtual reality. Almost every study has concluded that egocentric depth is underestimated in virtual environments for medium-field distances. A number of reasons for this have been proposed and tested such as the mechanical properties of the HMD (Willemsen et al. [99]), the field-of-view of the display (Knapp and Loomis [47], Creem-Regehr et al. [19]), a mismatch between the real world experiment site and its virtual representation, and the quality of the scene graphics (Willemsen et al. [101], Interrante et al. [41]). In summery, the exact cause of the depth underestimation in virtual environment is still unclear.

There are relatively very few depth perception studies for near-field distances in virtual reality. The VR depth perception studies have focused primarily on the effects of various depth cues and judgment techniques. Naceri et al. [66] studied the effects of environment fidelity (rich, reduced cues, poor) and apparent size (constant, co-varied) of virtual objects on depth judgments in near-field virtual reality. They used a collimated HMD, and a visually open-loop pointing task to judge egocentric distance to the target objects. Observers were more accurate for the rich cue condition as compared to the reduced cue and poor cue conditions. The manipulation in the target's size, however, led to individual differences among subjects. Napieralski et al. [68] studied near-field depth perception in real and virtual environments using both verbal and physical reach responses. Observers saw a virtual environment through a collimated head-mounted display. They found distance underestimation in both real and virtual conditions; the underestimation was less in the virtual environment as compared to the real environment.

The studies discussed above used collimated head-mounted displays, with a focus requirement of infinity. For near-field distances, a collimated display dissociates eye convergence and accommodation, which is usually coupled for normal real world viewing (see Figure 2.15). Bingham et al. [10] used blind reaching to study the effect of accommodation on egocentric distance estimation in a virtual reality environment. For the collimated condition, they found overestimation in the virtual environment relative to actual environment. They attributed this difference to accommodation being entailed farther than the virtual image in the collimated display. Interestingly, when a -$2D$ lens was used to reduce the focal distance, the overestimation reduced by half.

Anderson et al. [1] studied the effect of collimation on depth perception for near-field distances of $3-195$ cm by judging exocentric distance between two virtual objects. They found that judged depth was greater for the collimated condition than for the non-collimated condition. Also, as the simulated distance between the observer and the depth interval increased, the perceived extent of the depth interval decreased. They also found that there was a constant increase in depth with collimation for all of the studied depth intervals. This suggests that the space rescaling was uniform in the case of collimation.

3.3 Depth Perception in Augmented Reality

Very few depth perception studies have been conducted in augmented reality at medium field distances (e.g., Jones et al. [42], Swan et al. [84]). These studies found that egocentric depth is underestimated in augmented environments for medium-field distances. The exact cause of the depth underestimation in augmented environment is still unclear.

To the best of our knowledge, to date the published work examining near-field AR depth judgments can be found in only five papers (Rolland et al. [77, 78], Ellis and Menges [26], McCandless et al. [56], Singh et al. [83]). Rolland et al. [77] examined depth judgments of real and virtual objects at distances of 80 to 120 cm, using a forced-choice task. They found that the depth of virtual objects was overestimated at the tested distances. They attributed this overestimation to a number of factors: the fixed inter-pupillary distance of the display, optical distortion, illumination, and collimated display. Rolland et al. [78] then ran additional subjects with an improved display with a focal depth of 80 cm. They examined depth judgments of virtual objects at the focal depth distance of 80 cm, and compared forced-choice and perceptual matching tasks. They found improved depth accuracy and no consistent depth judgment biases.

Ellis[1] and Menges [26] studied depth perception at near-field distances, in order to develop design guidelines for augmented reality display systems. Their experiment involved measuring the effects of convergence, accommodation, observer age, viewing condition (monocular, biocular[2] stereo, binocular[3] stereo), and the presence of an occluding surface (the x-ray vision condition) on depth judgments of a virtual object using a closed-loop perceptual matching technique. The task was to match the depth of a small light on an adjustable arm to the depth of a virtual pyramid. They found that monocular viewing degraded the observer performance for depth judgments, and that most of the localization

[1]Stephen R. Ellis, NASA Ames, is a member of author's graduate committee, and has been a collaborator in the development of the experiments discussed in this dissertation.

[2]Biocular stereo: In this two identical images are presented to the eyes, but offset to indicate constant disparity.

[3]Binocular stereo: In this two stereoscopic images, which are generated from the perspective of each eye, are presented to the eyes.

errors occurred when the physical surface was closer than the virtual object. They attributed these errors to the phenomenon of proximal vergence: they found that introducing a physical surface in front of the virtual object caused a relative forward movement in the localization of the virtual object. They measured the eye convergence before and after introducing the physical surface using nonius lines, and found a small relative convergence of the eyes caused by proximal vergence. Then, they cut a hole in the occluding surface so that the virtual object was visible through the occluder. In this condition, the depth judgment bias towards the observer was reduced by a significant amount. This phenomenon appears to strengthen the proximal vergence explanation.

McCandless et al. [56] used the same experimental setup and task as Ellis and Menges [26] to additionally study motion parallax and AR system latency in monocular viewing conditions; they found that depth judgment errors increased systematically with increasing distance and latency.

3.4 Master Thesis

The work in this dissertation stems from the author's Master Thesis, titled "Near-Field Depth Perception in See-Through Augmented Reality" [82]. The goal of that work was to study depth perception in near-field augmented reality. As part of the thesis, a depth perception measurement apparatus was engineered and related calibration and measuring techniques were developed for collecting depth judgments (see Figure 3.1). To test this apparatus, a depth perception experiment was conducted, which compared *matching*, a closed-loop task for measuring depth judgments, with *reaching* task, a visually open-loop task for measuring depth judgments. The experiment also studied the x-ray vision condition by measuring the effect of a highly salient occluding surface appearing behind, coincident with, and in front of a virtual object. Additionally, this experiment also studied the relationship between dark vergence and depth perception. The major findings from this experiment were published in Singh et al. [83].

The experiment described in Singh et al. [83] was the first experiment of its kind to directly compare matching with reaching as a depth judgment protocol in augmented reality. This experiment replicated the apparatus and closed-loop task described in Ellis and Menges [26], along with the additional condition of reaching. These two depth judgment protocols were nested with two occluder conditions: *absent* and *present*. This nesting resulted into four main conditions. The scene was presented through a collimated NVIS nVisor ST60 optical see-through display. The virtual stimulus target was an inverted white, wireframe pyramid, with a base of 10 cm and height of 5 cm rotating at 4 revolutions per minute. It was randomly scaled from 70% to 130% of its actual size to avoid presenting relative size as a depth cue.

The main results from this experiment are shown in Figure 3.2. A major finding was that there was a general trend of a linear increase in error with increasing stimulus distance for all conditions. This trend was more notable when the occluder was not present, i.e. in conditions 1 and 3. This effect suggests that the convergence-accommodation mismatch present in the HMD might have biased the observer to converge farther away for father distances. When observers had the occluder (a real object) in view, i.e. in conditions 2 and

Tabletop Apparatus

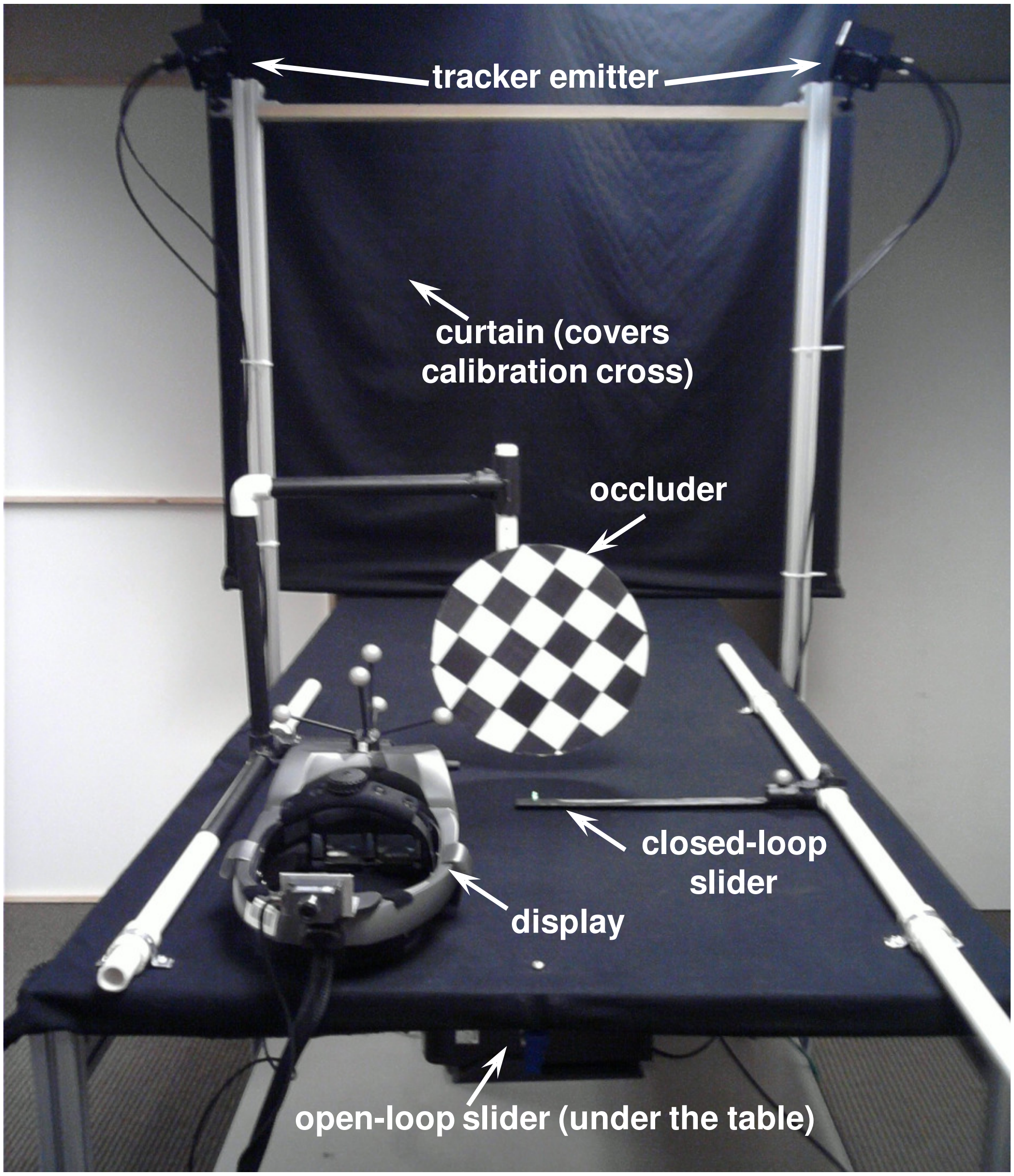


Figure 3.1

Table-top based experimental apparatus.

4, it disrupted the linear pattern, probably by biasing convergence towards the occluding

dista

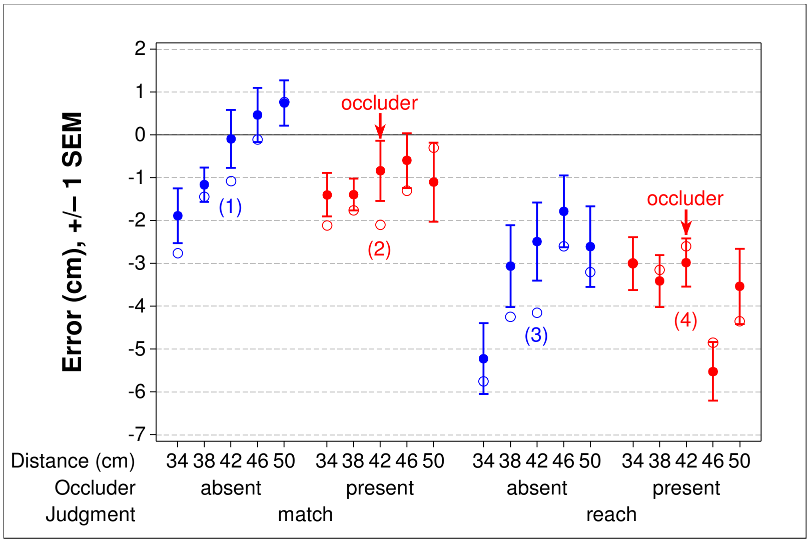


Figure 3.2

The mean error for the depth judgments.

The effect of the occluder was more prominent in the blind-reaching condition, especially for the distance of 46 cm. This result is consistent with an observation made by both us and Ellis and Menges [26]: when a virtual object is initially located in front of a physical object, and the physical object is slowly moved towards the observer, at first the virtual object appears to be "pushed" closer to the observer by the physical object; this is likely happening for the object at 46 cm in condition 4. At some point, however, the virtual object suddenly appears to "fall back" behind the physical object; this is likely happening for the object at 50 cm. This imparts a strong sense of transparency to the physical object. This effect is easy to see in an AR system using one's hand.

Finally, we found a general trend of greater underestimation for the reaching judgments relative to the matching judgments. This effect was expected, as there were fewer depth cues available in case of reaching than matching.

CHAPTER 4

EQUIPMENT & EXPERIMENTAL SETUP

4.1 A.R.T. GmbH ART TrackPack position tracking system

The *ART TrackPack* is a position tracking system (see Figure 4.1) with 6 degrees of freedom tracking capability. It is a passive outside-in tracker that uses a combination of retro-reflective spheres and tracking cameras to track the position and orientation of the target object. The tracking cameras consist of an infra-red light emitter and receiver, and are attached at a fixed location in space.

Figure 4.1

A.R.T. GmbH ARTtrack position tracking system.

A rigid configuration of the retro-reflective spheres is attached to the target to be tracked. During tracking, the cameras emit infra-red light, which is reflected off the spheres towards the same direction it came from. This reflected light is captured by the cameras and from this data the position and orientation of the target object is calculated by triangulation. Multiple target objects can be tracked at the same time by attaching different configurations of spheres to the objects. During preliminary testing, the accuracy and precision of the tracker was found to be better than 1 mm in the tracked volume.

4.2 Display Technologies

4.2.1 NVIS nVisor ST60 Optical See-Through HMD

Figure 4.2

NVIS nVisor ST60.

The first two experiments discussed in this document were conducted using an NVIS nVisor ST60, which is an Optical See-Through Display (see Figure 4.2) . This HMD is capable of displaying 3D stereo scenes at a resolution of 1280×1024 per eye. Its display area has 100% overlap with a 60° diagonal field-of-view. It supports variable inter-pupillary distance (IPD) within the range of 53 mm to 73 mm, by adjusting the left and right eye-pieces independently. The optical elements of the HMD are collimated and present the scene at the eye accommodation of ∞.

4.2.2 Haploscope

Historically, head-mounted displays (HMDs) and heads-up displays (HUDs) were mainly used for pilot training and other medium and large field applications, which primarily used pictorial cues for depth estimation. These displays, which are usually located a few centimeters in front of the eyes to a few meters away, can inadvertently provide accommodation information for medium and far-field applications. When looking through a display at closer distances, the eyes focus at the distance to the display due to the accommodation sensitivity. If the display is sufficiently close it gives the illusion of flatness, where all objects can appear flat, even though they are intended to be rendered at different depths with correct pictorial depth cues. This happens because at closer distances, the visual system starts incorporating depth information from the accommodation cue. As a result, all parts of the scene appear to be at the same distance. This lack of differential accommodation implies that the scene is flat and close rather than extended to medium or far field distances.

To solve this problem, the images are presented with collimated light in some displays that sets the accommodative demand at infinity and thus removes accommodation as a flatness cue (Nagata [67], Anderson et al. [1]). Because accommodation is mostly effective at closer distances and differential accommodation does not occur at farther distances, displays that are focused at infinity do not use accommodation information while estimating depth. This technique works well for medium to far-field distances; however, for near-field distances it causes the problem of convergence-accommodation mismatch that affects the way a scene is perceived.

At near-field distances, accommodation changes to bring an object in focus are very large (1 diopter to 10 diopters to switch focus from 1 meter to 10 cm in front of the eyes). Ideally, HMDs should provide these changes in accommodative demand while presenting objects at different distances. However, most of the current commercially available displays can not present virtual objects at their correct accommodative demand due to fixed optical properties. In these displays, the focal distance is fixed at a particular distance, and their inability to present virtual objects at correct accommodative demand is a major design issue in VR/AR displays. A major design limitation of the NVIS nVisor ST60 HMD (which is used in Experiments I and II of this dissertation) is that it is a collimated display and graphics can only be presented at the accommodation demand of infinity. As previously discussed in the related work section (see section 3.2), the convergence-accommodation mismatch causes perceptual problems when judging distance at closer distances.

For Experiments III, IV, and V, we needed to present AR objects with equal convergence and accommodation demands. We designed a *haploscope*, which is a display with switchable focus that can change accommodative demand based on the distance to target objects. The design of our haploscope was loosely based on the one explained in Rolland et al. [77]. Our major requirement was that the haploscope should provide optical see-through augmented reality functionality by enabling a direct view of real world along with virtual objects.

The haploscope (see Figure 4.3) consists of two image-generator assemblies, one each for left and right eye, along with a mechanism to mount these assemblies on. A critical requirement for the haploscope was that it should be rigid enough to keep the equipment stable and provide immunity to accidental bumps and environmental factors. At the same time, it should be flexible enough to incorporate variations in the inter-pupillary distance (IPD) of different subjects as well as the changes in the accommodative demand for different target distances.

To accomplish this, we used three Newport Precision optical rails to build the haploscope. A 24 inch optical rail was used as a mounting base for the haploscope, and two 12 inch optical rails were used as bases for the image generator assemblies for the left and right eyes. The longer optical rail was bolted firmly on to the optical bench of the table-top apparatus described in Singh [82]. The two smaller optical rails were mounted on to this rail by using two 3 inch rail carriers. These rail carriers could slide left and right on the optical rail to change the distance between left and right assemblies to match the IPD of various subjects.

Haploscope III, IV, V

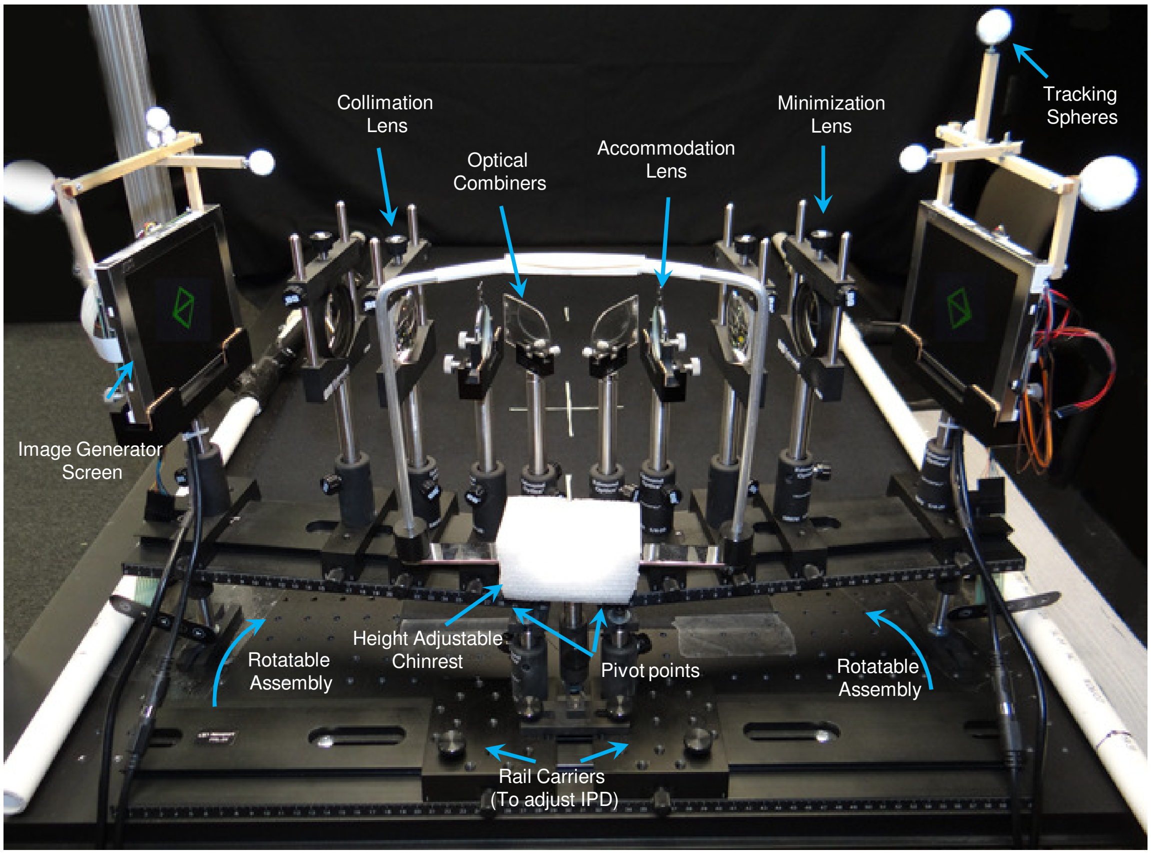


Figure 4.3

Augmented reality haploscope.

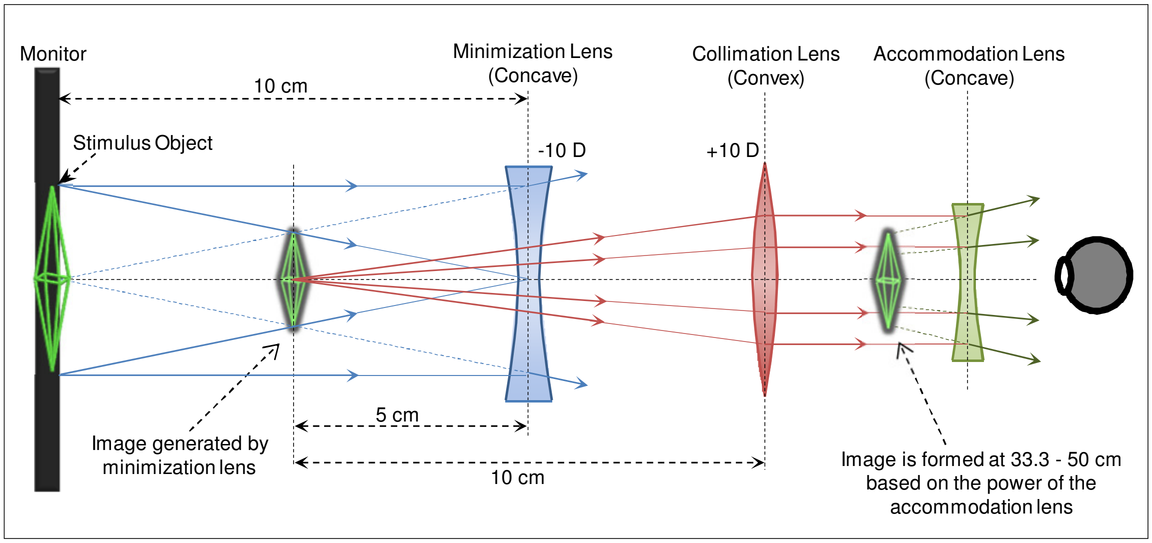


Figure 4.4

Ray diagram of image-generator assembly.

A simple image-generator assembly consists of a display screen and a collimation lens. For display screens, we used two off-the-shelf Accelevision 6 inches (15.2 cm) TFT color LCD monitors (9.2 cm high and 12 cm wide). For image collimation, we used two plano-convex lenses (Techspec PCX Lens, 50.0 mm Diameter $\times$ 100.0 mm FL, 1064 nm V-Coat acquired from Edmund Optics Inc.). After experimenting with the optics, we found that even though these lenses can collimate the incoming light, they were unable to capture the entire display area of the monitors, and therefore the usable display area that was visible through the collimation lens was very small. We decided to resize the image coming from the monitors by introducing two minimization lenses in between the monitors and the collimation lenses (see Figure 4.4). Therefore, instead of collimating the monitor screen, the collimation lenses were now collimating a minified image of the monitor image as presented by the minimization lenses. We used two plano-concave Lenses (TechSpec PCV

50 mm Diameter, -100 FL, VIS 0, Inked acquired from Edmund Optics Inc.) as minimization lenses. All of these lenses were held in place by English bar-type holders also acquired from Edmund Optics Inc. A ray diagram of the minimization and collimation in image-generator assembly is shown in Figure 4.4

When we actually started building the haploscope, we found the problem of optical distortion. It has been found that optical distortion can change the presented position of virtual objects and therefore could result into erroneous depth judgments (Rolland et al. [77], Wann et al. [96]). There were three different approaches to design the haploscope based on the effect of optical distortion on the target presentation. Figure 4.5 shows the first approach, which is appropriate when the available lens has minor or no distortion. In this case the entire structure of the haploscope is rigid (as in an HMD) and the eye convergence is provided using binocular disparity by generating a stereoscopic scene on the screens based on the position of the left and right eyes. With no lens distortion, this configuration can show the scene at the correct distance from the observer. However in our case, the lens distortion caused the virtual objects to appear at a distance closer than the intended distance as shown in Figure 4.5. The dotted-purple lines show the intended fixation point for the virtual object, and the dotted-black lines show the actually displayed fixation point.

The second approach (see Figure 4.6) is based on the relationship between distortion and the distance from the center of the lens. The distortion is smaller near the center of the lens as compared to the edge of the lens. If an object is projected near the center of the lens system, an acceptable amount of distortion can be achieved. To accomplish this, the left

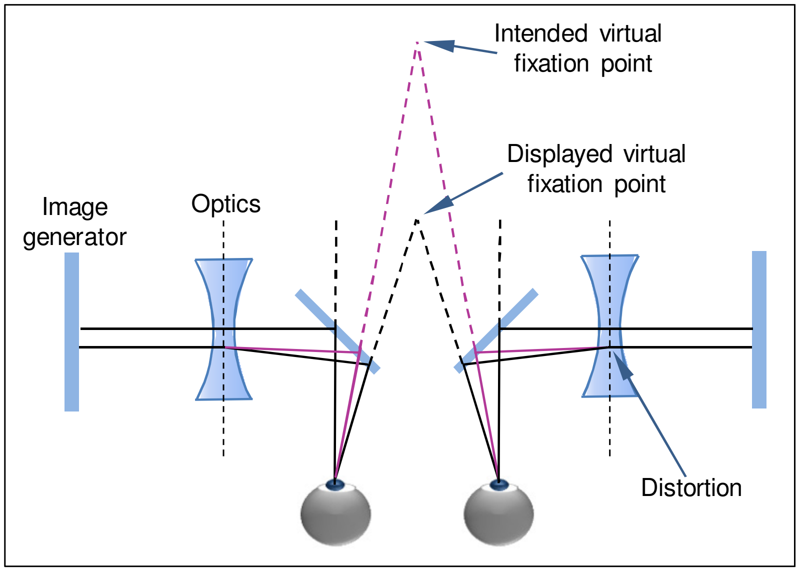
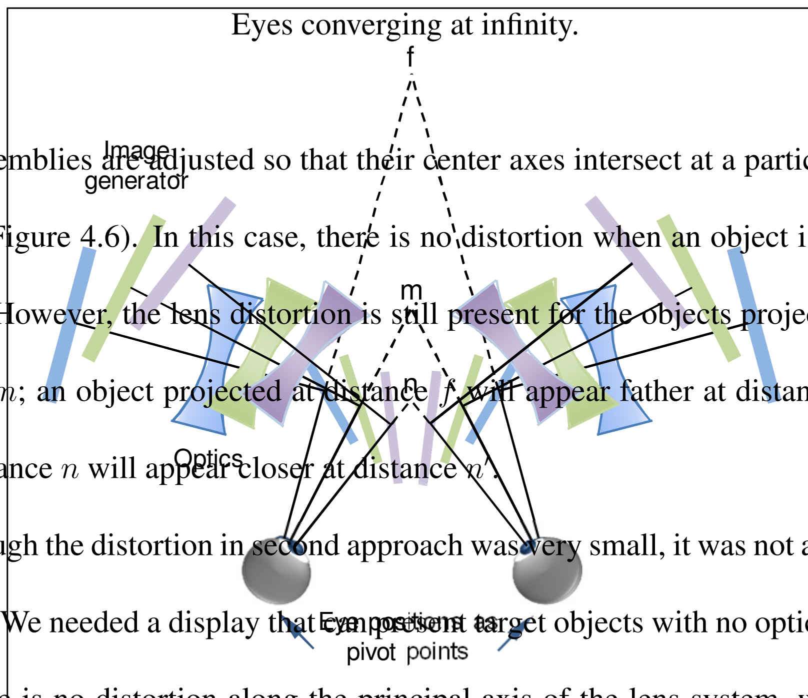
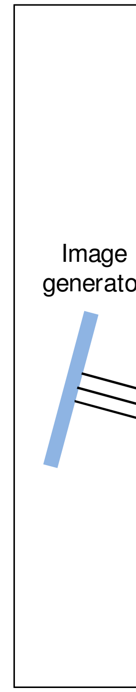


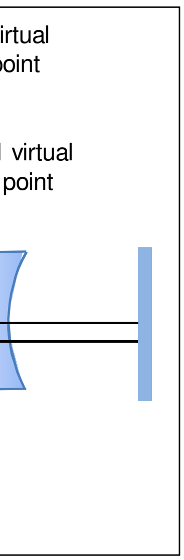
Figure 4.5





and right assemblies are adjusted so that their center axes intersect at a particular distance (point m in Figure 4.6). In this case, there is no distortion when an object is projected at distance m. However, the lens distortion is still present for the objects projected closer or farther than m; an object projected at distance f will appear father at distance f', and an object at distance n will appear closer at distance n'.

Even though the distortion in second approach was very small, it was not acceptable for our purpose. We needed a display that can present target objects with no optical distortion. Because there is no distortion along the principal axis of the lens system, we decided to render the target object at the center of the lens system. This technique would solve the problem of optical distortion, and present the virtual objects with the correct accommoda-



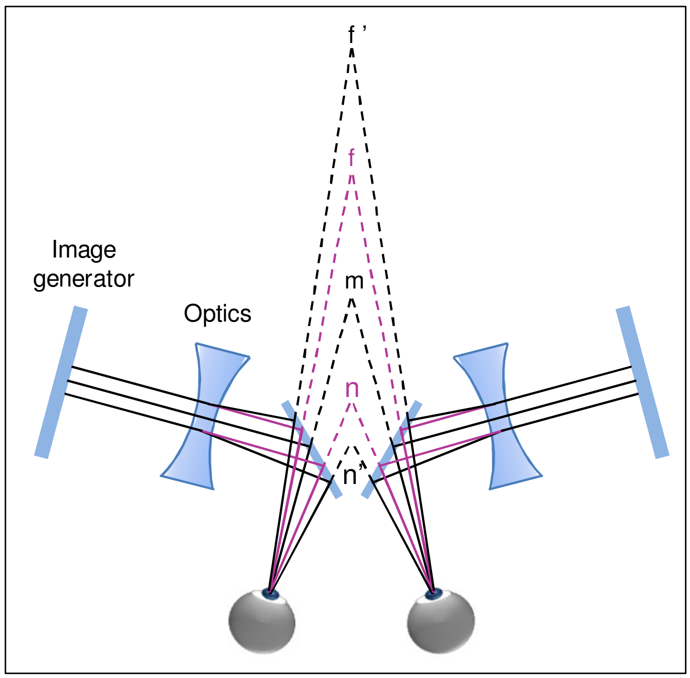
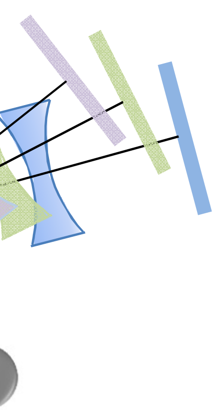
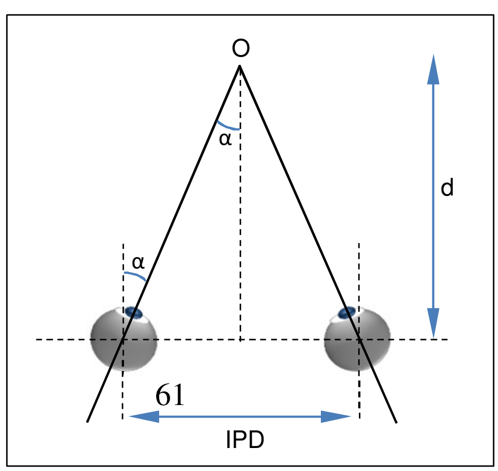


Figure 4.6

Eyes converging at midpoint.





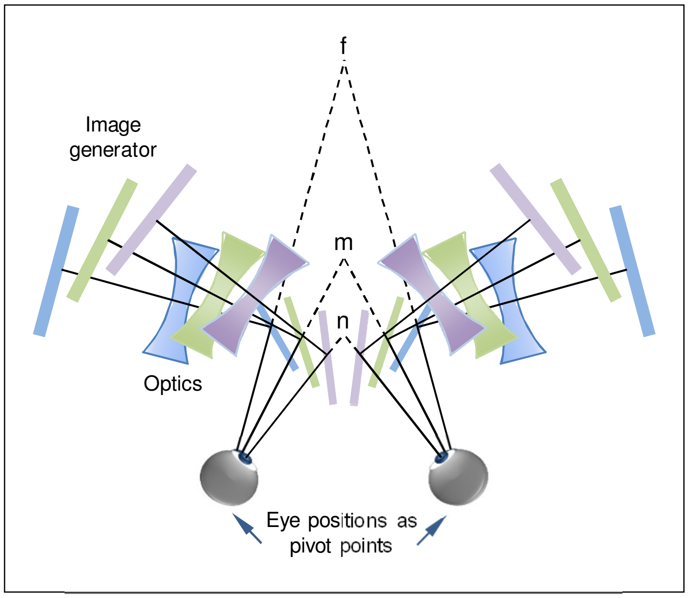
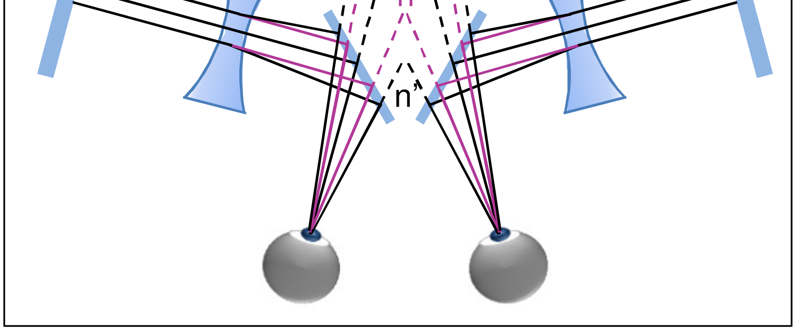


Figure 4.7

Eyes converging dynamically at selected distances.

In traditional stereoscopic systems (e.g. HMD), a scene is rendered stereoscopically based on the eye positions of an observer to provide binocular disparity. Because of this binocular disparity, viewing closer objects requires inward convergence of the eyes and farther objects requires outward convergence. In our system, rendering the target object by binocular disparity meant again running into the problem of optical distortion, as shown in Figure 4.5. Instead, we decided to provide the required disparity in the same way as the eyes provide during their normal operation; by rotating the entire left and right assemblies so that the principal axes of both assemblies intersect at the target distance. As shown



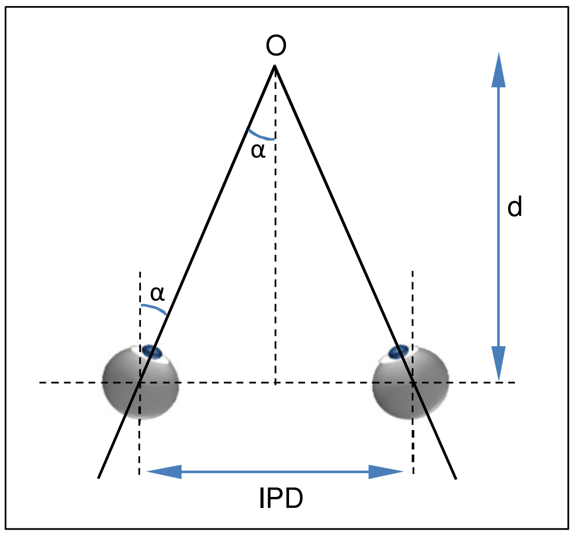


Figure 4.8

Eye convergence by an angle α.

To display the target object at a specific distance, the assemblies were rotated at the convergence angle that matches the presented object. This angle was calculated based on the inter-pupillary distance (IPD) of the current participant, and the distance to the object.

As shown in Figure 4.8, if d is the distance to the presented object O, the convergence angle α required by the left and right eye to look at the object is

$$\alpha = tan^{-1}\left(\frac{IPD}{2d}\right) \tag{4.1}$$

It was essential to rotate the assemblies at a high precision and accuracy to present target objects at their correct locations. We used the ART TrackPack position tracking system to measure the rotation of the assemblies by attaching retro-reflective markers to the image generator screens as shown in Figure 4.3. This system has a rotation accuracy of one-hundredth of a degree. While running subjects, the assemblies were rotated and the accommodation lenses were changed for each trial to present the target object with the desired accommodative and convergence demands.

4.2.3 ASD FieldSpec Pro Spectroradiometer

A spectroradiometer is an instrument that measures the reflection and transmission properties of a surface by providing spectral power distributions over the entire elecromagnetic spectrum originating from or reflected off of that surface.

We used a FieldSpec Pro spectroradiometer by ASD (see Figure 4.9) to measure the luminance of the target object. This device provided the spectral power distribution and radiance of the target object over all wavelengths, including visible and non-visible wavelengths. To calculate the luminance (the part of the measured radiance perceived by the human eye), we used the photopic luminosity function (Equation 2.5, Page 38), to calculate the weighted integration of radiance over the visible spectrum.

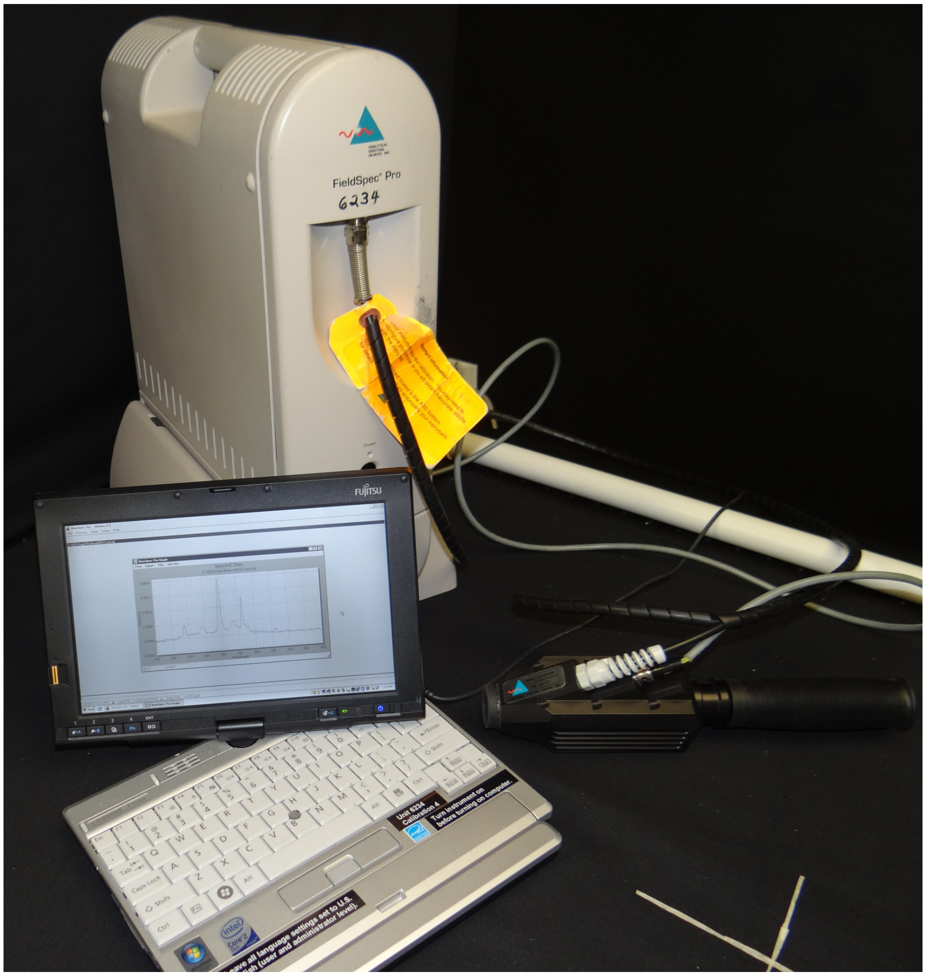


Figure 4.9

ASD FieldSpec Pro Spectroradiometer.

CHAPTER 5

EXPERIMENTAL PROCEDURES

5.1 HMD Calibration

Experiments I and II of this dissertation were conducted using the NVisor ST60 HMD. To present the stimulus object in a way so that it is correctly observed by the participants, it is necessary that the HMD is fitted correctly on their heads. We used the technique described in Jones et al. [42] to fit the display on the participant's head. This technique was originally developed for medium-field distances for a standing participant looking down a hallway; the technique's parameters are the participants eye height, inter-pupillary distance, and the distance from the participant's eyes to a calibration cross mounted at the participant's eye height at the end of the hallway. We modified this technique to work with our setup for near-field distances, using an eye-height of 3.5 cm above the table surface for Experiment I, and an eye-height of 24.5 cm for Experiment II. A white cross was drawn at eye-height on a black cardboard surface that was mounted 100 cm from the participant.

The original technique described in Jones et al. [42] was developed for medium-field distances. However, we discovered a problem with this technique when using it at near-field distances. This problem results in a lateral separation of up to 2 cm between the real and virtual objects. At medium field distances a 2 cm separation is relatively small and can be ignored. However, 2 cm is a large error at near-field distances.

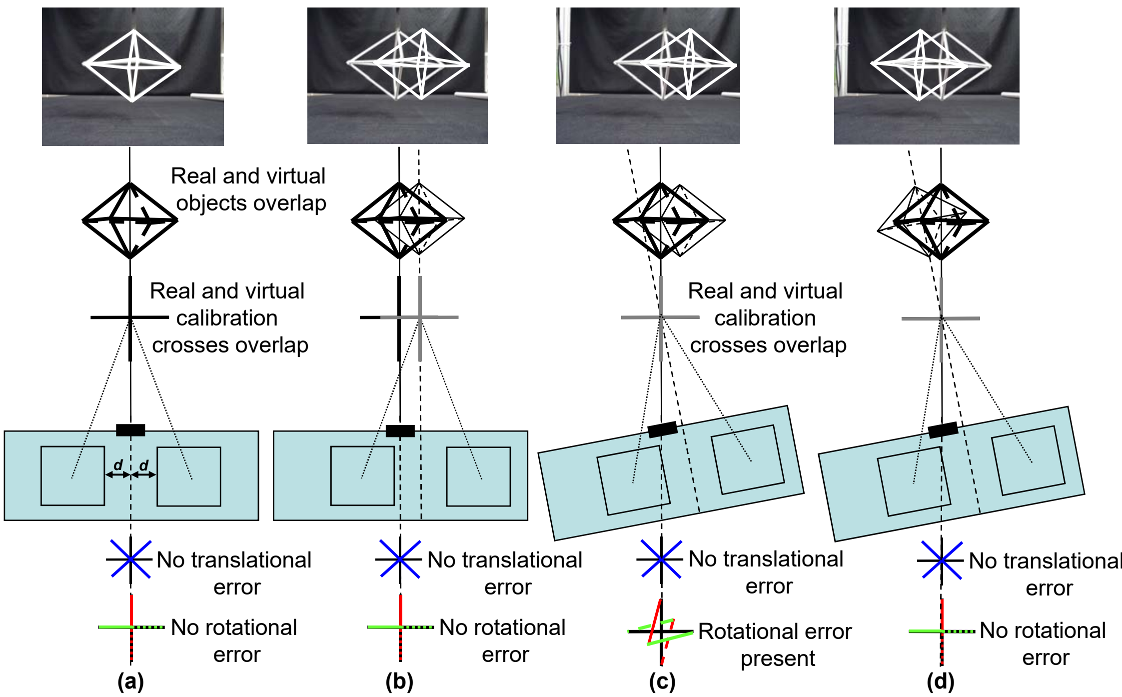


Figure 5.1

HMD calibration problem.

Figure 5.1 illustrates the problem. This problem arises when the HMD is shifted to one side of the participant's head. Figure 5.1(a) shows the ideal case, where the HMD eyepieces are at equal distance from the center of the HMD. When the participant aligns everything (the real and virtual calibration crosses, the translational crosshair, and the rotational crosshair; see Jones et al. [42]) the system is calibrated, and therefore real and virtual objects overlap. Figure 5.1(b) shows what happens when the HMD's eyepieces are not centered. At the start, when the translational and rotational crosshairs are aligned, there can be a lateral separation of 2 cm between the real and virtual calibration crosses, because the virtual calibration cross is generated with respect to the center point. This results in shifting the whole scene towards the direction of misalignment. Figure 5.1(c) shows how participants compensate for this: they rotate their head to eliminate the lateral separation between the calibration crosses; this configuration does not result in any translation error, but it does generate rotational error that shows up as a misalignment in the rotational crosshair. Figure 5.1(d) shows what happens when participants correct this rotational error. Here both the translational crosshair and the rotational crosshair are properly aligned, but the real and virtual objects, which should be collocated, do not overlap.

Therefore, the calibration technique described by Jones et al. [42] only works when the eyepieces are centered in the display (Figure 5.1(a)). The problem described here was difficult to find because often the eyepieces are properly centered. However, Figure 5.2(a) shows what happens when the HMD is worn while shifted to one side of the head. The first step of the calibration procedure described by Jones et al. [42] is to monocularly adjust concentric circles so that equal amounts of the outermost circle are visible; this ensures that

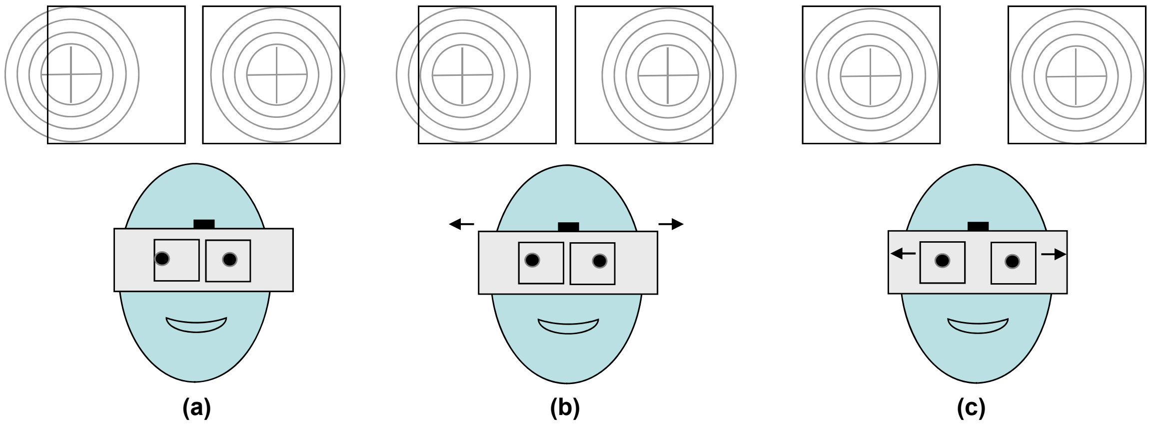


Figure 5.2

HMD calibration solution.

each of the participants eyes are looking down the middle of each monocles optical axis. In Figure 5.2(a) the participant would adjust the left monocle more than the right, resulting in the calibration problem described above.

Figure 5.2 describes the solution. First, both monocles are adjusted all the way inward. If the situation of Figure 5.2(a) occurs, the participant shifts the entire HMD left and right on the head until they have a symmetric view of each set of concentric circles (Figure 5.2(b)). This ensures that the HMD is centered on the head. Then, as shown in Figure 5.2(c), the participant adjusts each monocle until equal amounts of the concentric circles are visible. This solves the calibration problem described here.

5.2 Luminance Calculation

We used an ASD FieldSpec Pro spectroradiometer (Figure 4.9) to measure the radiance of the real target, the bright virtual target, and the dim virtual target. To measure the

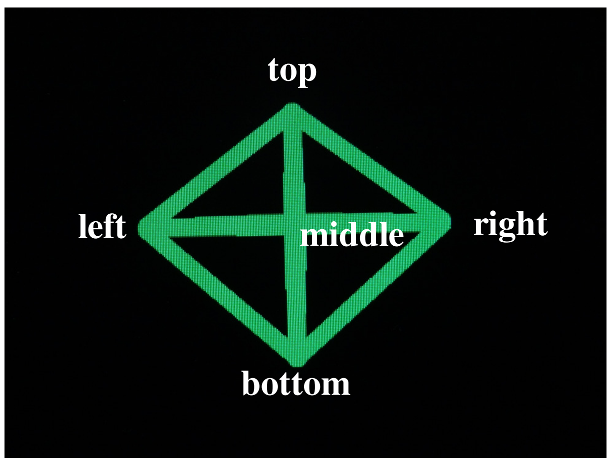


Figure 5.3

Luminance measurement locations for the virtual target.

The radiance values were then converted into luminance values according to equation 2.5 (Page 38). Figure 5.4 shows the luminance values for the three types of target object and five measurement positions. The filled circles show the median positions of the luminance for each target object.

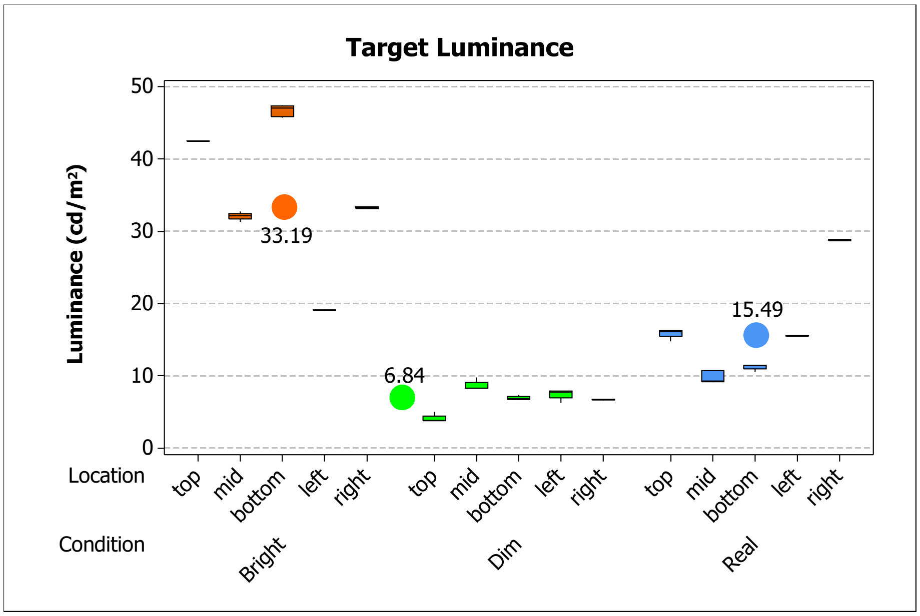


Figure 5.4

Luminance measurement locations for the target object.

CHAPTER 6

EXPERIMENTATION

The experiments in this dissertation studied depth perception in the near visual field for an augmented reality environment. The work stemmed from the author's masters thesis (Singh [82]) that studied the two complementary distance judgment protocols of perceptual matching and blind reaching, and found perceptual matching to be a more accurate distance measure than blind reaching. In this dissertation we evaluated various techniques to measure near-field depth perception in AR, and used these techniques to empirically study various factors that affect AR depth perception, specifically learning effects, accommodation, brightness, and age.

- **Experiment I:** ***Matching vs. Reaching***: This experiment compared the two complementary distance judgment protocols of perceptual matching and blind-reaching, using a physical pointer in a real, and in an AR environment. The primary result of this experiment is that participants were more accurate when using the perceptual matching technique than when using the blind reaching technique.

- **Experiment II:** ***Learning Effects***: This experiment investigated learning effects using a pretest, intervention, posttest experimental design. We found an effect of perceptual matching on the blind-reaching responses only in the real environment. This experiment also compared perceptual matching and blind-reaching but using a different setting with the observer's finger as the pointing object, to provide proprioceptive feedback and to better match previous work (Tresilian and Mon-Williams [88]). The results from this experiment also confirmed the validity of perceptual matching as an accurate distance measure.

- **Experiment III:** ***Accommodation***: This experiment used perceptual matching to study the effect of three accommodative conditions: collimated, consistent, and midpoint, on depth perception of younger participants. We found the distance responses in the consistent and midpoint conditions to be more accurate than in the collimated condition.

- **Experiment IV:** ***Brightness***: This experiment used perceptual matching to study the effect of brightness on depth perception by comparing distance judgments in two brightness conditions: bright, and dim. We found that the participants judged the objects in the bright condition closer than the objects in the dim condition.
- **Experiment V:** ***Age***: This experiment used perceptual matching to study the effect of age by testing the three accommodation conditions (collimated, consistent, midpoint) on older participants. We found that the older participants judged the distance more accurately than the younger participants.

The following sections will describe these experiments in detail:

6.1 Experiment I - Matching vs. Reaching

The main goal of Experiment I was to compare the two complementary distance judgment protocols of perceptual matching and blind-reaching at near-field distances in a real and in an AR environment. At near-field distances, most of the previous experiments have used perceptual matching depth judgments, where the depth of a target object is matched with a pointer object. For many imagined AR applications, perceptual matching has good ecological validity; for example, many AR-assisted medical procedures involve placing a medical instrument at a depth indicated by a virtual marker. However, many perceptual scientists do not consider perceptual matching to be an appropriate measure of depth perception, because it can only measure the depth of one object relative to that of another object (e.g., Bingham and Pagano [11]). These scientists have suggested blind reaching, where a participant indicates a distance by reaching with their unseen hand, as an alternative depth judgment measure that better reflects the perception of depth.

Motivated by these suggestions, we compared perceptual matching (called matching hereafter) and blind reaching (called reaching hereafter) in an AR environment in our previous near-field depth perception experiment (Singh et al. [83]). That experiment found that matching is a more accurate distance measure than reaching. However, because that experiment was conducted in an AR environment, a valid critique is that it only studied AR targets, but these virtual targets do not have a ground-truth location that can be objectively measured in the real world. Experiment I addressed this concern by comparing AR targets with real targets. Since the location of real targets can be objectively measured and hence can serve as ground truth, they will help ground the localization of AR targets.

We also found calibration effects over the first 5 trials in Singh et al. [83]. A valid critique about these calibration effects is that the experiment used a within-subjects, repeated-measures design, and such designs can have complicated asymmetric skill transfer or training effects across successive conditions. This suggests the possibility that experience in the earlier conditions could affect performance in the later conditions. This concern becomes even more prominent considering that reaching is not always stable over time (e.g., Mon-Williams and Bingham [59]), and therefore the reaching protocol combined with the within-subjects design complicates the interpretation of the calibration effects seen in Singh et al. [83]. Experiment I addressed this concern by replicating the experiment in Singh et al. [83] using a between-subjects design. Therefore, if additional calibration effects are found, the between-subjects design will confirm that they do not result from asymmetric skill transfer. However, if no calibration effects are found, it will indicate that the calibration effects seen in Singh et al. [83] resulted from that experiment's within-subjects design. Experiment I tested the following hypotheses:

- **Hypothesis 1.1:** ***Matching more accurate than reaching***: This hypothesis compared matching and reaching tasks. We hypothesized that matching will give more accurate results as compared to reaching, because the matching task uses more depth cues (e.g., disparity) than blind reaching. Therefore, matching might serve as a ground truth or best-possible depth judgment method, even if the results only measure relative depth perception. The degree of divergence between reaching and matching will serve as an important measure of the usefulness of reaching as a perceptual measure of depth perception.
- **Hypothesis 1.2:** ***Real objects localized more accurately than AR objects***: This hypothesis compared localization of AR objects and real objects. Since all depth cues are consistent for real objects, but not for AR objects, we hypothesized that distance judgments for the real objects will be more accurate as compared to the AR objects.
- **Hypothesis 1.3:** ***No calibration effects for between-subjects design***: This hypothesis tested if the calibration effects seen in Singh et al. [83] were present because of the within-subject design of the experiment. We hypothesized that because of the between-subject experimental design, Experiment I will not have any calibration due to training effects.

6.1.1 Experimental Setup

We used the same apparatus as described in Singh et al. [83], with modifications to support a real stimulus object. Figure 6.1 shows a side-view diagram of the table and depicts the depth judgment tasks and environments, and Figure 6.2 shows an annotated photograph of the apparatus.

We used closed-loop matching and open-loop blind reaching tasks as described in Singh et al. [83], with modifications to support similar arm movements for both of the tasks. In that study, each task required a different motor action. The matching task involved reaching to the right side of the table, while the reaching task involved reaching underneath the table-top. In the current experiment, we connected both pointers to a single

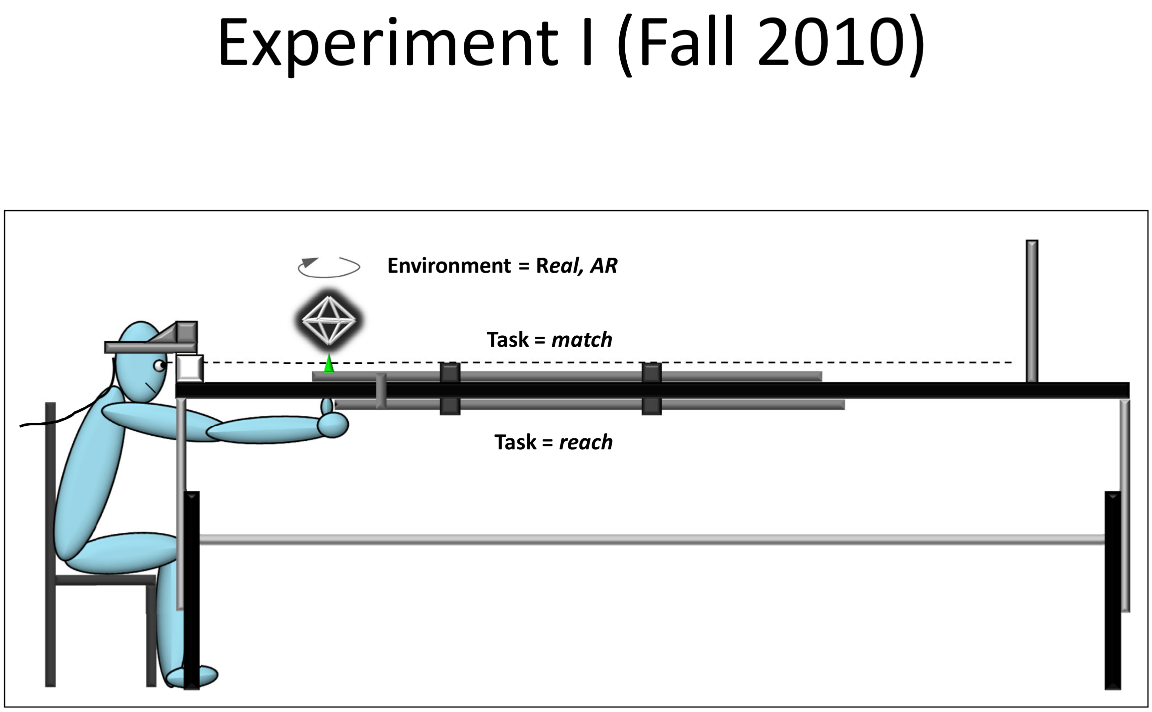


Figure 6.1

Experiment I: Side view diagram of the experimental table.

In the real environment, participants saw a slowly rotating (4 rpm) white wireframe diamond shape with a 10 cm base and 10 cm height (See Figure 6.1). As shown in Figure 6.2, we attached the target object to a PVC arm connected to a pipe that ran through two collars mounted on the left-hand side of the table. With this mounting, the target object could be positioned either in or out of the participant's field of view. In the real environment, we positioned the physical target in the field of view at a variety of distances in front of the participant. In the augmented reality environment, we displayed a virtual rendering of the

Experiment I (Fall 2010)

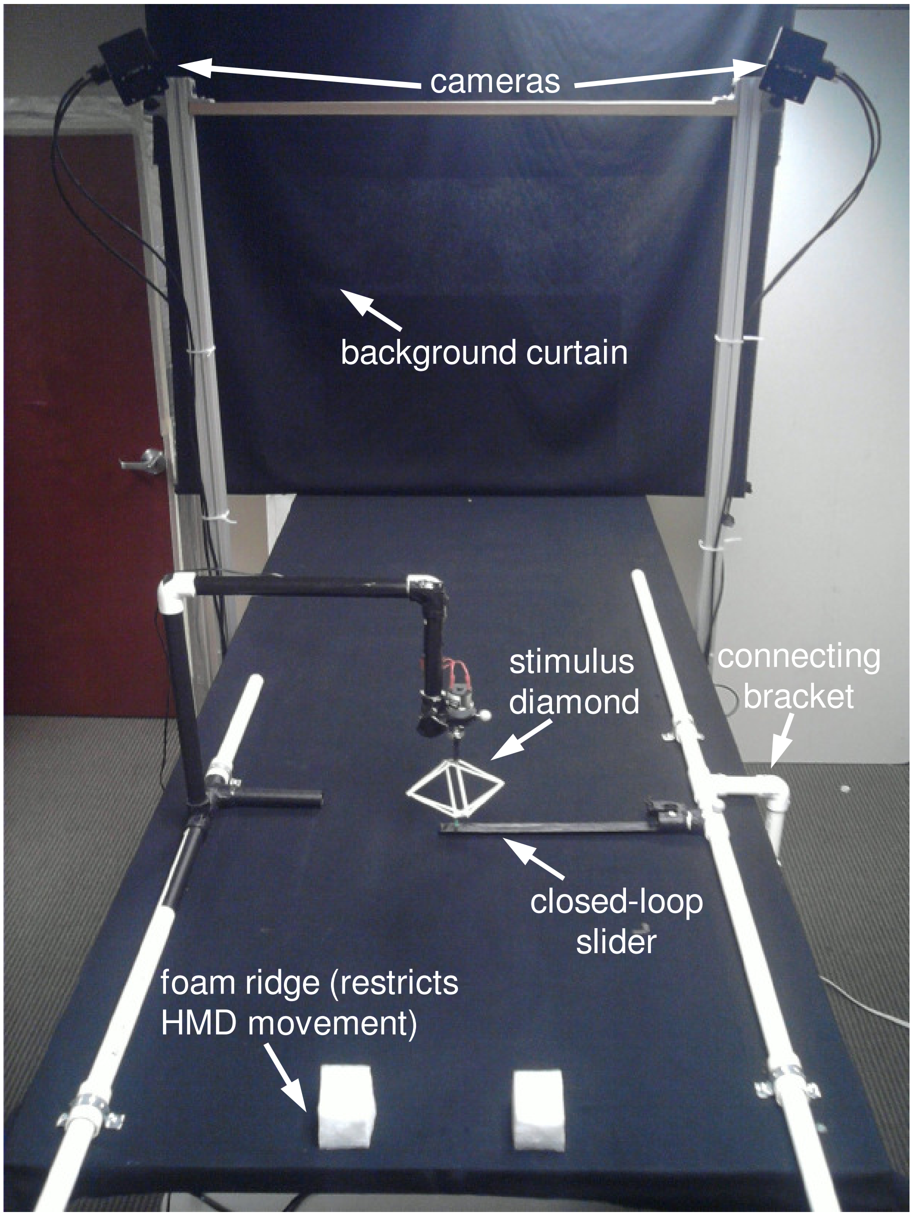


Figure 6.2

Experiment I: Annotated photograph of the experimental table.

same target object. In both environments participants viewed the target object through our AR head-mounted display, an nVisor ST model by NVIS, Inc. In the real environment the HMD did not show any graphics. Participants viewed the target object against a black curtain that hung 220 cm away (see Figure 6.2).

6.1.2 Experimental Variables and Design

Table 6.1 shows the independent and dependent variables of Experiment I. In this study, we recruited 41 participants from a population of university students, faculty, and staff. The participants ranged in age from 18 to 27; the mean age was 20.1, and 23 were female and 18 male. Seven participants were paid $12 an hour, and the rest received course credit. We did not analyze data of one participant as they observed incorrect disparity resulting from an accidental flip of the left and right eye. As indicated in Table 6.1, we analyzed the data from the remaining 40 participants.

Table 6.1

Experiment I: Independent and Dependent Variables

INDEPENDENT VARIABLES			
Participant	40		(random variable)
Environment	2	between	AR, Real
Judgment	2	between	Match, Reach
Distance	5	within	34, 38, 42, 46, 50 cm
Repetition	6	within	1, 2, 3, 4, 5, 6
DEPENDENT VARIABLES			
Judged Distance			in cm
Error	Judged Distance − Actual Distance		

The participants performed two kinds of the depth judgment tasks: matching and reaching, in two kinds of environments: real and augmented reality (AR). For the matching task, the participants used their dominant hand to manipulate the slider located underneath the table; this slider adjusted a small LED mounted on the closed-loop pointer above the table. The participants' task was to align the bottom of the target object with the top of the LED, which matched the depth of the bottom of the target object. For the reaching task, we disconnected the closed-loop pointer arm (with LED) from the slider. The participants adjusted the slider underneath the table until they believed that their thumb was located directly underneath the tip of the target object. Because the participants rested the front of the AR display on the tabletop, they could not see their hand, and so this task was performed blind.

We presented the target objects at 5 different distances from the participant: 34, 38, 42, 46, and 50 cm. At 34 cm distance, the target object subtended a maximum of 16.4° of visual angle, while at 50 cm distance, the target object subtended a minimum of 11.3°. Participants saw 6 repetitions of each distance.

As shown in Table 6.1, the primary dependent variable was judged distance, which we measured using either the matching or the reaching distance judgment. We also calculated $error = judged\ distance - actual\ distance$. An $error \sim 0\ cm$ indicates an accurately judged distance; an $error > 0\ cm$ indicates an overestimated distance; and an $error < 0\ cm$ indicates an underestimated distance.

We used a mixed design with the most important independent variables (*judgment*, *environment*) varying between the participants, resulting in four main conditions: (1) matching, real (2) matching, AR (3) reaching, real, and (4) reaching, AR. The presentation order of these four main conditions varied in a round-robin fashion. For each main condition, our control program generated a list of 5 (*distance*) $\times$ 6 (*repetition*) = 30 *trials*. The program then randomly permuted the presentation order of the distances, with the restriction that the same distance could not be presented twice in a row. We collected 1200 data points [40 *participants* $\times$ 5 *distances* $\times$ 6 *repetitions*].

6.1.3 Procedure

Each participant filled out a standard informed consent form, a simulator sickness survey, and a demographic survey. Next, each participant performed a stereo acuity test, which all participants passed. After this, we measured the participant's inter-pupillary distance (IPD) while they converged at a distance of 40 cm; and we entered this IPD value into the scene generator software. Next, we described the depth judgment task to the participant, and they practiced the task three times. During practice, the participant did not wear the display and used the real target object for both the real and AR conditions. Next, we fitted the display on the participant's head, and calibrated it using the calibration technique described in our previous study (Singh et al. [83]). In addition, while designing this study we discovered a problem with the calibration tecnique; in section 5.1 we describe the problem in detail, as well as the additional calibration step that we developed to solve it.

After calibration, the participants looked through the approximate optical center of each of the display eyepieces, without any translational or rotational errors. In addition,

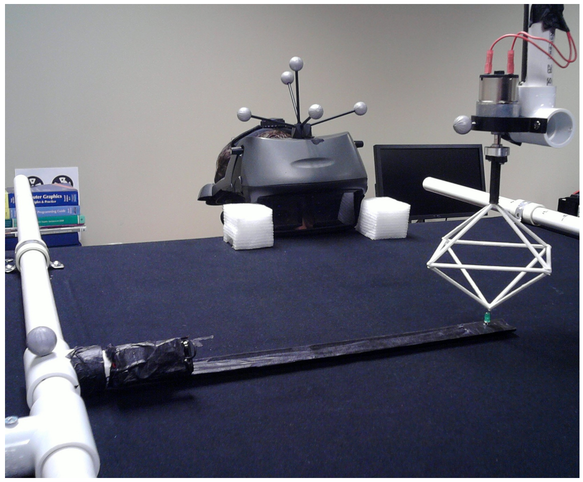


Figure 6.3

Experiment I: A participant after calibration.

the front of the HMD rested on the table, and any lateral movement was restricted by the two ridges on either side of the display (see Figure 6.3). This was done because the field of view of the HMD is narrow, and restricting the movement of the HMD equalized the field-of-view between the AR and real conditions. This setting also allowed us to avoid any depth information from motion parallax resulting from head movements. In addition, most of the HMD weight was held on the table, which helped us eliminate the effect of HMD weight on depth perception (Willemsen et al. [100]).

The participants next completed one of the four main *judgment* $\times$ *environment* conditions of 30 trials. We allowed the participants to take a break at any point during the experiment; if a participant took a break we re-calibrated the display. At the beginning of the trials, we adjusted the slider so that it was as close to the participant as possible. For the first trial, the participant slid the pointer from this position to indicate the depth of the target object. At the beginning of every trial, the participant closed their eyes and we presented the target object. Then, we asked the participants to open their eyes and adjust the pointer. When the participant completed the task, they placed their hand in their lap, and closed their eyes. When we saw the participant do this, we blanked the display, recorded the judgment, and then triggered the next trial. We ask the participants to be as accurate as possible, and gave them no time limit for performing the trials. After completing the experiment, participants filled out a simulator sickness exit survey, and then discussed the experiment with us.

6.1.4 Results

We analyzed the judged distance results from $N = 1200$ data values. All analysis were conducted with *Error* = *Judged Distance*$-$*Actual Distance*. Figure 6.4 shows the main results as judged distance versus actual distance and Figure 6.5 shows this same data plotted as error. We conducted an ANOVA analysis of this main experimental design, and found a significant 3-way interaction between judgment, environment, and distance ($F(4,144) = 2.7,\ p < .031$). Included in this interaction are a main effect of judgment ($F(1,36) = 46.7,\ p < .001$), and a main effect of distance ($F(4,144) = 6.8,\ p < .001$). There is no effect of environment ($F(1,36) = 0.01,\ p = .75$). This is a complex interaction; there is a different pattern of means by distance for each of the four main conditions.

For (1) the matching, real condition, the judged distance is extremely accurate, with a *mean error* of $+0.14$ *cm*. For (2) the matching, AR condition, the judged distance shows a linear increase with increasing distance, beginning with an *error* of $+0.52$ *cm* and pro-

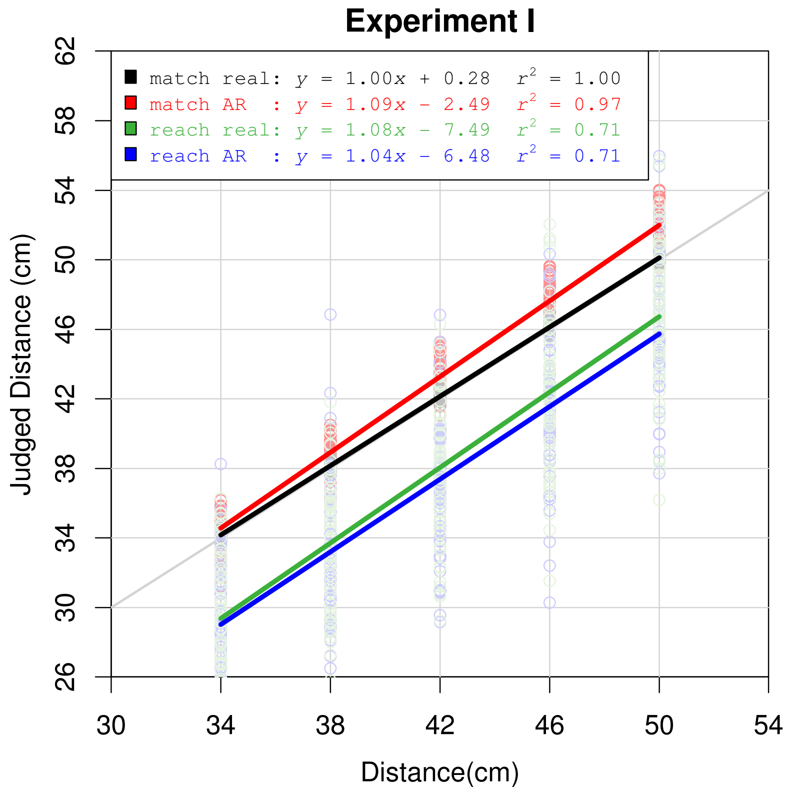


Figure 6.4

Experiment I: Mean judged distance versus the actual distance ($N = 1200$).

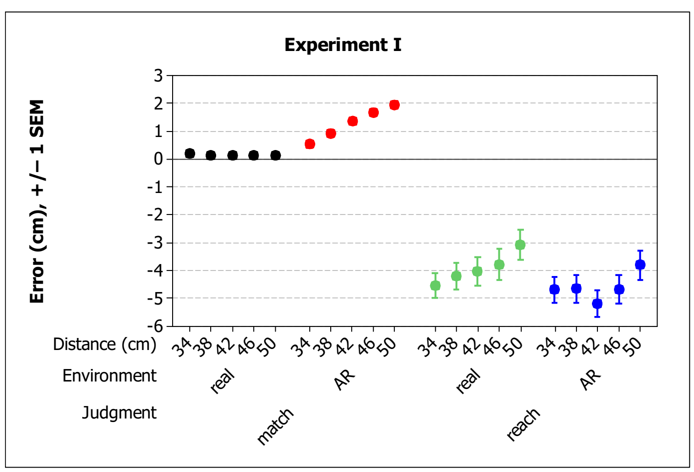


Figure 6.5

Experiment I: Mean error for the overall depth judgments ($N = 1200$).

gressing to an *error* of $+1.88$ *cm*. In this condition, there is a strong effect of distance on error (F(4,36) = 72.5, $p < .001$), and a Ryan REGWQ post-hoc homogeneous subset test indicates that every mean is significantly different at $p \leq .05$. The judged distance means are well-described by a linear model with $m = 1.09$ and $b = -2.49$ (r^2 = 97%); this model indicates that at these near-field distances, for every additional centimeter of distance, participants localized the target an additional 0.9 mm farther away than its veridical location. For (3) the reaching, real condition, the judged distance means are underestimated, with a *mean error* of -3.94 *cm*. For (4) the reaching, AR condition, the judged distances are underestimated, with a *mean error* of -4.62 *cm*.

We further analyzed the reaching judgments for the real and the AR environments, and found no significant effect of environment on error (F(1,18) = .23, $p = .64$). The reaching judgment responses did not differ between the real and the AR environments. Additionally, we analyzed the data for effects of trials and repetition, but did not find any significant effects in any condition.

6.1.5 Discussion

In the real environment, the participants were extremely accurate at judging distance in the matching task. These results are consistent with the real life experiences of adult humans who can reach and grasp objects in the real world with high accuracy and precision (Mon-Williams and Bingham [59]). The visual and haptic feedback that is available in everyday interactions with real objects calibrates our internal sense of depth, and thus increases the accuracy of any future interactions with real objects (Bingham and Pagano [11], Bingham et al. [10]). Even though the real target object was viewed through the HMD, the matching responses were extremely accurate, indicating that the reduced field-of-view of the HMD did not affect participants' ability to judge distance in the real environment. One potential reason of this accurate performance is that the participants might have primarily used disparity cue to judge distance, which remained unaffected by any changes in the field-of-view. Overall, these results indicate that matching is indeed a very accurate and precise measure of distance at near-field distances.

The matching responses in AR were overestimated and the error increased linearly with increasing distance. Despite matching being an accurate measure of distance judgments, the matching responses in the AR environment are very different from the matching responses in the real environment. This divergence of the matching responses in the AR environment from the best possible matching responses in the real environment indicates perceptual differences in these two environments. A potential source of the differences in matching responses is the way AR objects are rendered in the head mounted display. In particular, we used a collimated HMD that rendered the target objects at an accommodative demand of infinity, resulting in a dissociation between convergence and accommodation.

It has been well established that when convergence and accommodation are separated to attend different distances, the normal coupling of the convergence-accommodation reflex pulls convergence in the direction of accommodation, and results in misjudgment of distance towards the accommodation distance (Bingham et al. [10], Ellis and Menges [26], Mon-Williams and Tresilian [61], Peli [73], Wann et al. [96]). In our experiment, when matching AR targets, the participants increasingly overestimated distances, from $\sim$0.5 to

$\sim$2 cm (m =1.09), with increasing distance. This increase is consistent with the collimated optics of the head-mounted display driving the participants' vergence angle outward by a constant amount (e.g., Mon-Williams and Tresilian [62]).

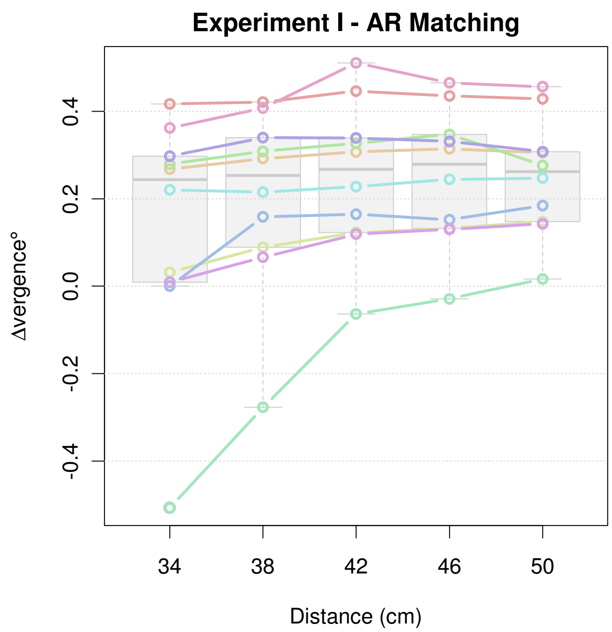


Figure 6.6

Experiment I: The change in vergence angle with distance.

Figure 6.6 illustrates this hypothesis in terms of the change in vergence angle at each distance for each of the 10 participants in the AR matching condition. The change in vergence angle ΔV is calculated as $\Delta V = \alpha - \beta$ (see Figure 6.7), where α is the angle of binocular parallax at the presented distance, and β is the angle of binocular parallax at the judged distance. Figure 6.6 shows that for 9 of the 10 participants ΔV changes less than 0.2°, and the median line seen in the underprinted boxplot changes less than 0.1° with

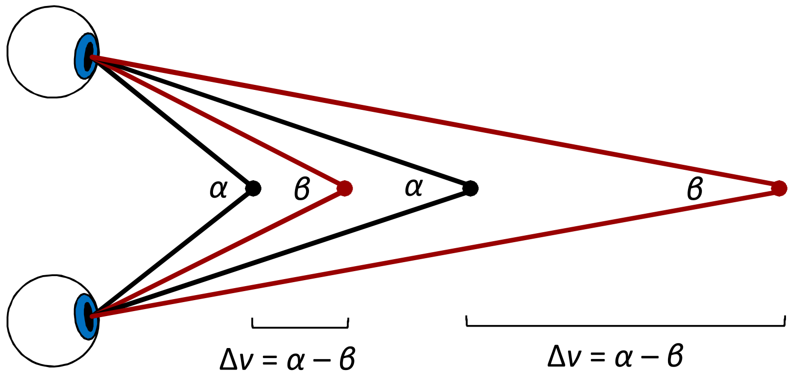


Figure 6.7

Disparity changes for AR matching.

changing distance. These angular changes are small, and indicate that the change in vergence angle is relatively constant with increasing distance. This strongly suggests that the collimated display optics are driving the vergence angle outwards by a constant amount, which causes increasingly overestimated depth judgments with increasing depth. These matching results suggest that, for near-field distances, accommodative demand needs to more closely match actual distances for accurate AR depth judgments; collimated optics cause overestimated depth judgments at near-field distances. Since the distance judgments were accurate for the real objects but not for the AR objects, our Hypothesis 1.2 is confirmed; real and AR objects are perceived differently.

The reaching responses in the real, as well as, in the AR conditions were underestimated, and were not significantly different. In the real environment, previous depth perception studies using reaching tasks have found a variety of results; overestimation of the egocentric distance (Foley [30], Foley and Held [32]), and very accurate distance judgments (Tresilian and Mon-Williams [87, 62], Von Hofsten and Rosblad [92], Van Beers et al. [90]). For our reaching task, we aimed to replicate the findings of Tresilian and Mon-Williams [87, 62]. They studied egocentric depth perception in the near-field and found very accurate distance responses as shown in Figure 6.8. In our experiment, the reaching responses were underestimated for both the real (*error* = -3.94 *cm*), as well as in the AR environment (*error* = -4.62 *cm*) (see Figure 6.4). These results failed to match the accur

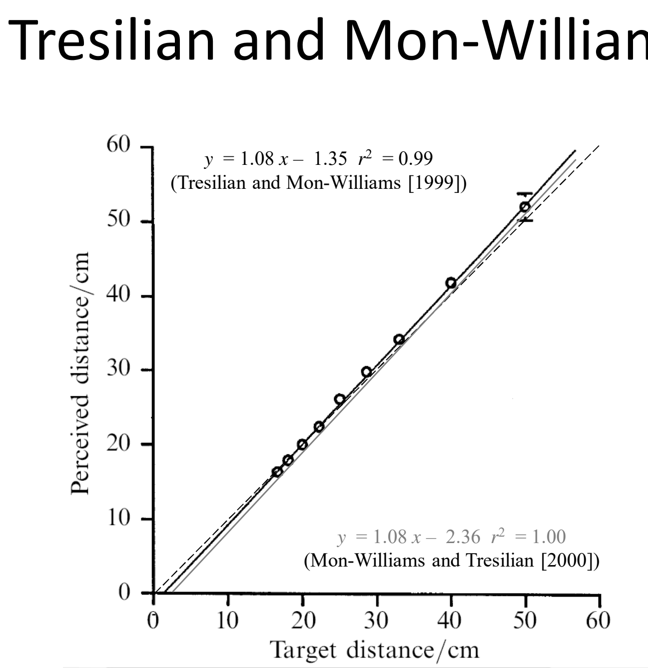


Figure 6.8

Experiment I: The results from Tresilian and Mon-Williams [88] (black) and Mon-Williams and Tresilians [62] (grey).

One potential reason for our results being different than that of Mon-Williams and Tresilian's results is the different biomechanical actions of the participants for the reaching tasks. In Mon-Williams and Tresilian [61, 62], the participants reached to the right side of the target object, and used the unseen index finger of their right hand to indicate the distance. In our experiment, the participants reached underneath a table-top to match their thumb to the distance of a target above the table. Also, the vertical distance between the target object and the thumb was about 7 cm in our experiment, which was larger than the one in Mon-Williams and Tresilian's experiment (they did not quantitatively report the distance between the target object and the index finger but mentioned it as "in the immediate vicinity"). We tested the hypothesis that these task differences explain why we obtained such different results than Mon-Williams and Tresilian in our Experiment II reported later in this document.

Another interesting finding of our results is that we did not find any significant difference in the reaching responses of the real and the AR environment; both were underestimated. A possible explanation of the reaching results is that the perceived distance is the same in the real and AR environments, and that the reaching response was similarly miscalibrated in the two environments.

These result can be indicative of the instability that is inherent in reaching tasks. It has been found that depth judgments using visually open-loop tasks (such as reaching) are susceptible to drift in the absence of some corrective feedback (Bingham and Pagano [11], Wann and Ibrahim [95]). The lack of feedback destabilizes the system, and therefore, the effects of the studied visual perturbations become difficult to resolve. Instead, Bingham and Pagano [11] advocate using perception-action tasks with feedback to measure distance perception. They argue that this approach would help understand the effects of perturbed visual information independently from the instability problems associated with the motor control in visually guided tasks. In our reaching task, the difference between the mean errors of the real and the AR environment was 0.68 cm. Considering the variability in the reaching responses in both environments, this difference is trivial and non-significant. The lack of significant difference between the reaching responses of the real and the AR environment indicates the inability of the reaching task to detect the perceptual differences between the real and AR environments. However, another interpretation is that real and AR objects are perceived similarly, as demonstrated by reaching; but are not matched similarly because of the collimated display.

Overall, these results confirm our Hypothesis 1.1 as the matching responses were more accurate as compared to the reaching responses for both the real as well as the AR environment. These results also confirm our Hypothesis 1.2 that real objects are perceived more accurately than AR objects.

Unlike Singh et al. [83], we did not find any calibration effects in any of the conditions. The lack of calibration effects in Experiment I indicate the robustness of the between-subjects experimental design against training effects. These results also indicate that the calibration in Singh et al. [83] resulted from asymmetric skill transfer due to the within-subjects design of that experiment. This finding confirms Hypothesis 1.3 that the between-subjects design did not result into calibration effects. We can sum up the results for the hypotheses that we stated at the start:

- **Hypothesis 1.1** was confirmed by the results of Experiment I. The perceptual matching task was relatively more accurate than the blind reaching task for both real and AR environments.

- **Hypothesis 1.2** was confirmed by the results of Experiment I. The real objects were judged more accurately than the AR objects for the matching task, while not significantly different for the reaching task.
- **Hypothesis 1.3** was confirmed by Experiment I. We did not find any significant calibration or learning effects for any of the conditions. These results confirm that the between-subjects design of Experiment I did not result in calibration effects.

6.2 Experiment II - Training Effects

In Experiment I, we found matching to be a more accurate measure of depth perception than reaching. Consistent with past research, the matching responses were very accurate in the real environment (Bingham and Pagano [11], Bingham et al. [10]), while they were overestimated with a linear increase in error with distance in the AR environment (Singh et al. [83]). The reaching responses, on the other hand, did not show the accuracy seen in previous studies (e.g., Tresilian and Mon-Williams [88], Mon-Williams and Tresilian [61, 62]), and they were underestimated for both the AR and the real environment.

Because the accuracy of these previous results (see Figure 6.8) originally motivated us to study reaching as a depth judgment technique, in Experiment II we more closely replicated the apparatus and task of Tresilian and Mon-Williams [88]. Experiment II compared matching and reaching with a bio-mechanical action similar to Tresilian and Mon-Williams [88]. Instead of using a physical pointer, participants used the index finger of their right hand to judge distance; an action that provided consistent proprioceptive feedback.

Experiment I established that the calibration effects seen in Singh et al. [83] resulted from asymmetric skill transfer due to the within-subjects design of the experiment. The participants were calibrating or learning the task in one condition and using that knowledge in the other conditions. Experiment II studied this learning effect by employing a ***pretest, intervention, posttest*** experimental design. The pretest and posttest measured the depth judgments before and after an intervention, and any differences between pretest and posttest judgments would show up as the training effect resulting from exposure to the intervention.

This experimental design was motivated by AR surgical applications, where surgeons would be matching the depth of tools to virtual objects. We wanted to measure the effect of this kind of matching task on reaching judgments. We tested RMR (reaching as the pretest and posttest phases, and matching as the intervention phase) to measure the effect of matching on reaching. We expected that the matching would improve the reaching. We also tested MRM (matching as the pretest and posttest, and reaching as the intervention) so that initial matches (M in MRM) could be compared to initial reaches (R in RMR) and to complete the design. This experimental design also facilitated a direct comparison between Experiment I and Experiment II, by comparing the Experiment I results with the pretest results of Experiment II.

Experiment II also addressed individual differences among participants. In the near-field, the internal sense of depth is primarily established and calibrated by reaching and grasping real objects in everyday activities (Bingham and Pagano [11]), and therefore a normal individual's internal representation of distance is based on their arm length. Since every individual has a different arm length, a fixed set of distances may not be suitable for measuring depth perception in a variety of subjects. To measure depth perception in a normalized manner, Experiment II used a set of distances that were calculated based on the maximum reach of the participants.

Experiment II tested the following hypotheses:

- **Hypothesis 2.1:** ***Matching more accurate than reaching***: This hypothesis compared matching and reaching as distance measures. We hypothesized that matching will be more accurate as compared to reaching, because the matching task uses relatively more depth cues than reaching without visual feedback. We expected that the results will replicate the finding of Experiment I that matching is more accurate than reaching.
- **Hypothesis 2.2:** ***Reaching in Experiment II more accurate than Experiment I***: This hypothesis compared the effect of biomechanical action between Experiment I and II. We hypothesized that reaching responses will be more accurate in Experiment II as compared to Experiment I, because the reaching task in Experiment II has more intuitive proprioceptive feedback. However, we expected the matching responses of Experiment II would be similar to Experiment I, as they use similar visual depth cues despite different motor actions.
- **Hypothesis 2.3:** ***Calibration due to intervention***: This hypothesis tested the presence of calibration effects resulting from exposure to the intervention condition. We hypothesized that the intervention condition would affect the distance judgments by skill transfer, and there will be a difference between the pretest and posttest judgment responses for both matching and reaching.

6.2.1 Experimental Setup

We used the same apparatus as in Experiment I, with modifications to perform the judgment tasks with right hand instead of the physical pointer. Figure 6.9 shows a side-view diagram of the table and the distance judgment tasks and environments, and Figure 6.10 shows an annotated photograph of the apparatus.

In the real environment, the participants saw the same target object (a wireframe white diamond shape) as Experiment I. The apex of the target object pointed right and the participants could reach it with their right hand. The apex was at the eye level (25 cm above the

Experiment II (Spring 2011)

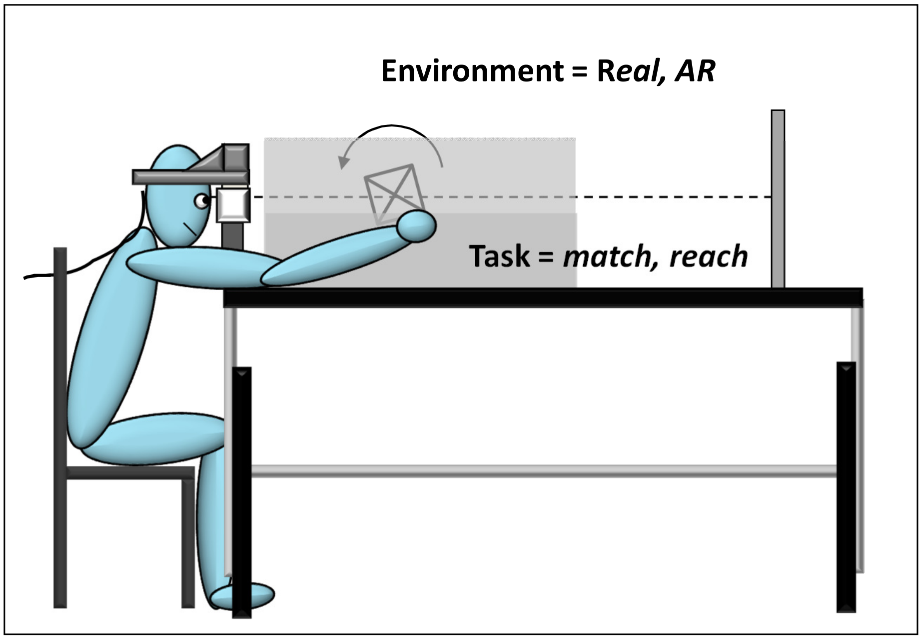


Figure 6.9

Experiment II: Side view diagram of the experimental table.

tabletop) and in the line of sight of the right eye of the participants. Similar to Experiment I, the target object could be positioned at variety of distances in front of the participants, and was seen against black cardboard positioned 72 cm from the participants (see Figure 6.10).

For the matching task, we used an occluding surface (0.5 cm $\times$ 23.7 cm $\times$ 61 cm) positioned 3 cm to the right of the target object. It stood upright by two supporting bars (38 cm high each) positioned at 9.5 cm and 71.5 cm from the front end of the table (see Figure 6.10). In the matching task (see Figure 6.11 (a)), participants judged the distance to the apex of the target object by resting their index finger on top of the occluding surface, while binocularly viewing the apex of the diamond, as well as the tip of their finger. Participants' responses were tracked by a sensor that they wore on their index finger. They

Experiment II (Spring 2011)

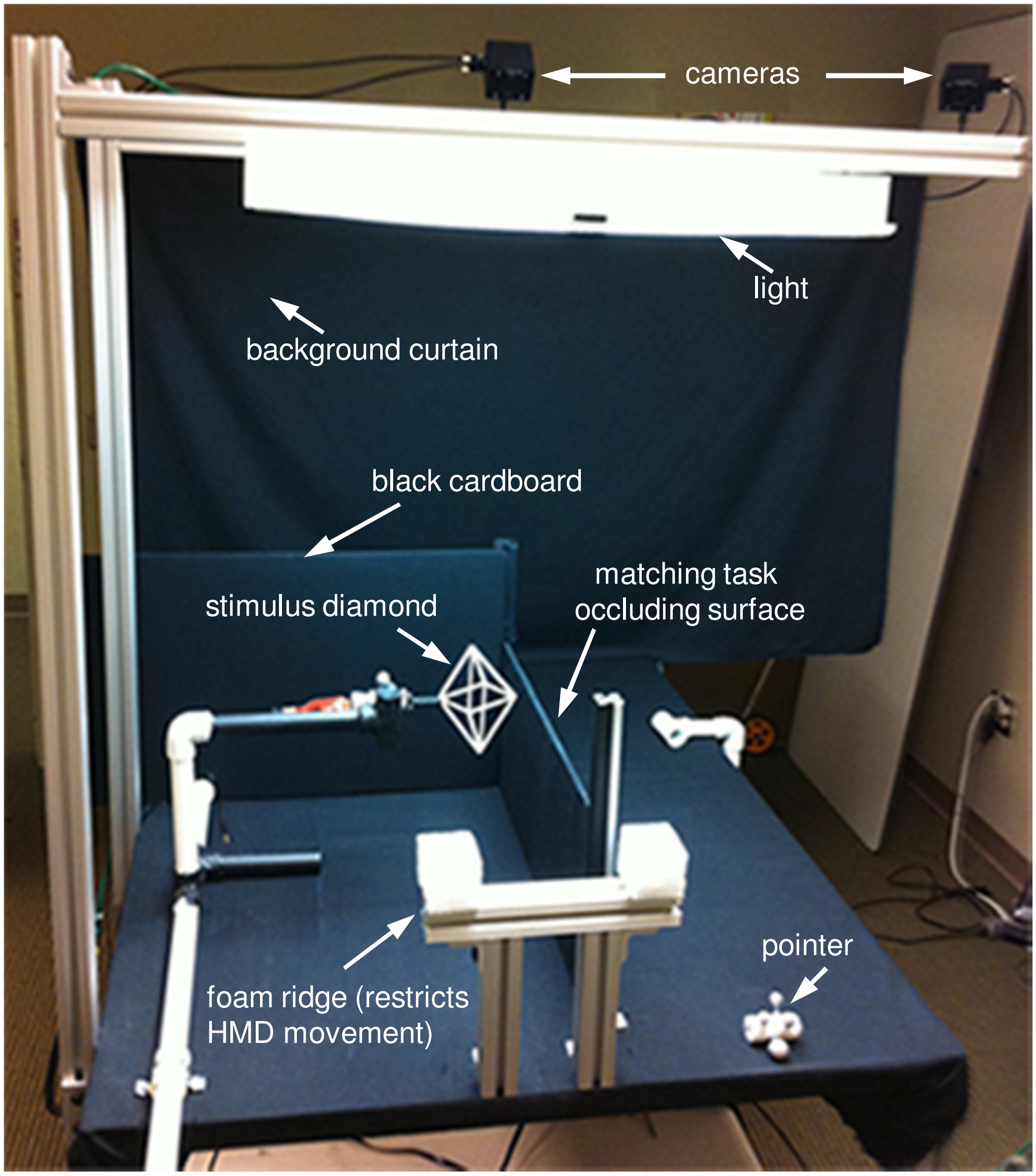


Figure 6.10

Experiment II: Annotated photograph of the experimental table.

(Equalized Photos)

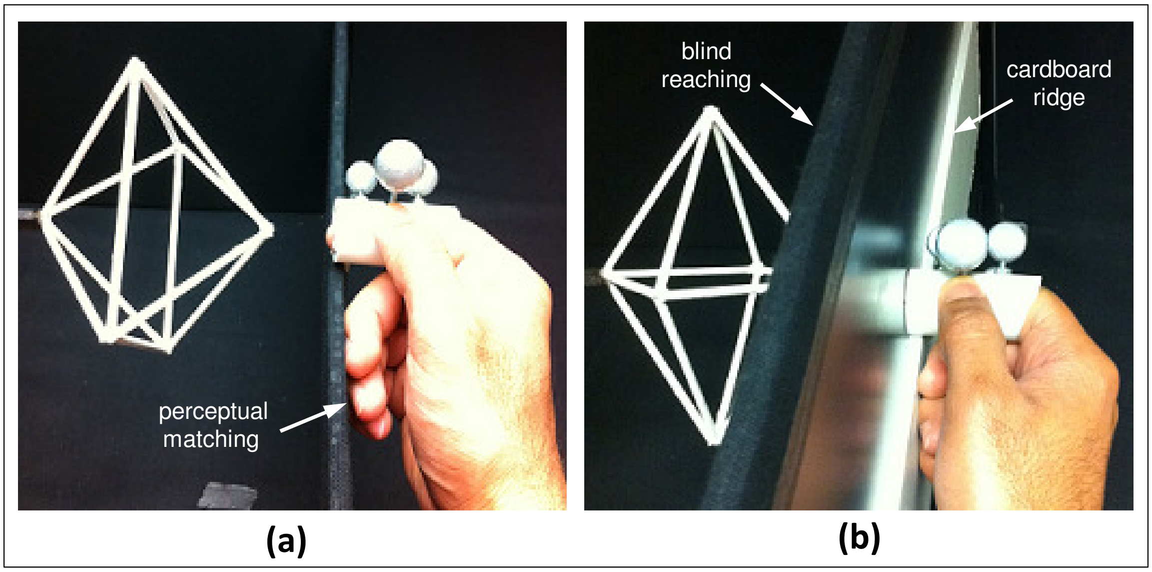


Figure 6.11

Experiment II: (a) Matching, and (b) reaching tasks.

were asked to match the tip of the nail of their index finger with the apex of the diamond, providing a specific point to match to another specific point, which facilitated fine-grain judgments.

For the reaching task (see Figure 6.11 (b)), a second occluding surface (0.5 cm $\times$ 38 cm $\times$ 61 cm) was positioned 3 cm to the right of the target object. The occluder was high enough to hide the right side view, and therefore participants could not see their hand during the reaching task. A cardboard ridge was attached to the right face of the occluder at 23.7 cm height (the same height as the matching task), and participants rested their index finger on the top edge of the strip while making reaching judgments. These two designs result in exactly the same biomechanical movement for the two tasks.

6.2.2 Experimental Variables and Design

Table 6.2 describes the experimental variables of Experiment II. In this study 42 participants were recruited from a population of university students. The participants ranged in age from 19 to 28; the mean age was 20.7, and 27 were male and 15 female. The interpupillary distance (IPD) ranged from 5.05 cm to 6.70 cm; the mean IPD was 5.90 cm. The maximum reach of the participants ranged from 43.3 cm to 64 cm; the mean maximum reach was 55.1 cm. Participants received course credit for their participation. We did not analyze data from two participants: the first participant had vision problems during the experiment, and the second participant did not receive the correct stimulus as the left HMD screen went blank in the middle of the experiment. As indicated in Table 6.2, we analyzed the data from the remaining 40 participants.

Table 6.2

Experiment II: Independent and Dependent Variables

INDEPENDENT VARIABLES			
Participant	40		(random variable)
Environment	2	between	AR, Real
Judgment	2	between	Match-Reach-Match (MRM) Reach-Match-Reach (RMR)
Distance	5	within	55, 63, 71, 79, 87% of Maximum Reach
Repetition	6	within	1, 2, 3, 4
DEPENDENT VARIABLES			
Judged Distance			in cm
Error			Judged Distance − Distance

The depth judgment tasks were arranged in two *judgment* conditions: MRM, and RMR. In MRM, participants performed matching task, then reaching task, and then matching task, while in RMR, participants performed reaching task, matching task and then reaching task. Participants performed the tasks in two kinds of environments, real and augmented reality (AR). In the real environment, a physical white diamond shape from Experiment I was presented as the target object, while in the AR environment the virtual rendering of the same object was presented. The target object was presented at 5 different distances calculated as 55, 63, 71, 79, and 87% of the maximum reach of the participant. The distances ranged from 23.8 cm to 55.7 cm for a maximum reach range of 43.3 to 64 cm. The participants saw 4 repetitions of each distance.

As shown in Table 6.2, the primary dependent variable was judged distance, which we measured using either the matching or the reaching depth judgment. We also calculated $error = judged\ distance - distance$.

We used a mixed design with *environment* and *judgment* varying between the participants resulting in four main conditions: (1) real, RMR (2) AR, RMR (3) real, MRM, and (4) AR, MRM. The presentation order of these four main conditions varied in a round-robin fashion. The distance and repetition varied within each one of the three judgment protocols per judgment condition per participant. For each judgment protocol, our control program generated 5 (*distance*) $\times$ 4 (*repetition*) = 20 *distances*. The program then randomly permuted the presentation order of the distances, with the restriction that the same distance could not be presented twice in a row. Therefore, every judgment condition consisted of

3 (*judgment protocol*) $\times$ 5 (*distance*) $\times$ 4 (*repetition*) = 60 *distances*. We collected 2400 data points [40 *participants* $\times$ 3 *judgment protocols* $\times$ 5 *distances* $\times$ 4 *repetitions*].

6.2.3 Procedure

At the start of the experiment, each participant filled out a standard informed consent form, a simulator sickness survey, and a demographic survey. Next, each participant performed a stereo acuity test, which all participants passed. Next, we measured the participant's inter-pupillary distance and entered this value in the control software. We then measured the participant's maximum reach. The participants wore a tracked sensor on the index finger of their right hand, and reached as far as they comfortably could while keeping their torso still. We recorded the position of their index finger. Based on the maximum reach distance, the control software generated a set of target distances. We described the judgment tasks to the participants and they practiced three times for each condition without wearing the HMD.

Next, the participant put on the HMD, and we performed our calibration procedure. The participants rested the front of the HMD on a platform, with an eye height of 23.7 cm off the table. Any lateral movement of the HMD was restricted using two ridges on either side of the display (see Figure 6.12) to prevent participants from seeing their finger while reaching, and to ensure that the target was in the field of view at all distances.

The participants next completed one of the four main *environment* $\times$ *judgment* conditions of 60 trials (20 trials each for pretest, intervention, and posttest). Participants were allowed to take a break at any point during the experiment; we recalibrated the display if

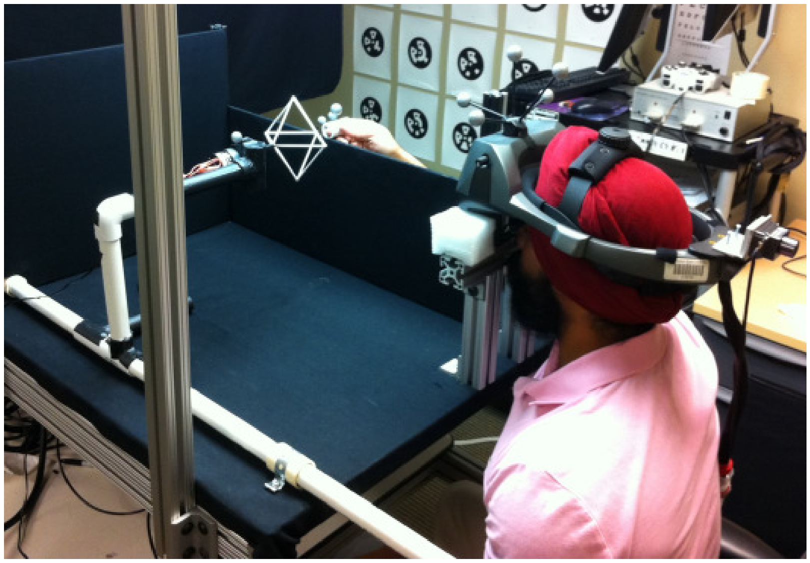


Figure 6.12

Experiment II: A participant after calibration.

participants took a break. At the beginning of every trial, the participants closed their eyes and we presented the target object. The participants opened their eyes and made the depth judgment with their index finger; we recorded the judgment position. The participants then closed their eyes and put their hand on the tabletop and the next trial began. We asked the participants to be as accurate as possible, without any time limit for performing the trials. After completing the experiment, participants filled out a simulation sickness exit survey, and then discussed the experiment with us.

6.2.4 Results

The analysis of Experiment II include two parts the pretest analysis (Section 6.2.4.1), and the overall analysis (Section 6.2.4.2). The pretest part of the experiment included data from the ”M” part for MRM participants and the ”R” part for RMR participants, and therefore provided between-subject data that was directly comparable to Experiment I.

6.2.4.1 Pretest Results

Figure 6.13 shows the pretest results as judged distance versus actual distance and Figure 6.14 shows this same data plotted as error. We conducted an ANOVA analysis on error. There are strong main effects of judgment (F(1,36) = 17.75, $p < .001$), environment (F(1,36) = 16.63, $p < .001$), and distance (F(4,144) = 13.67, $p < .001$). There is also a significant 2-way interaction between environment and distance (F(4,144) = 3.41, $p =$ .011).

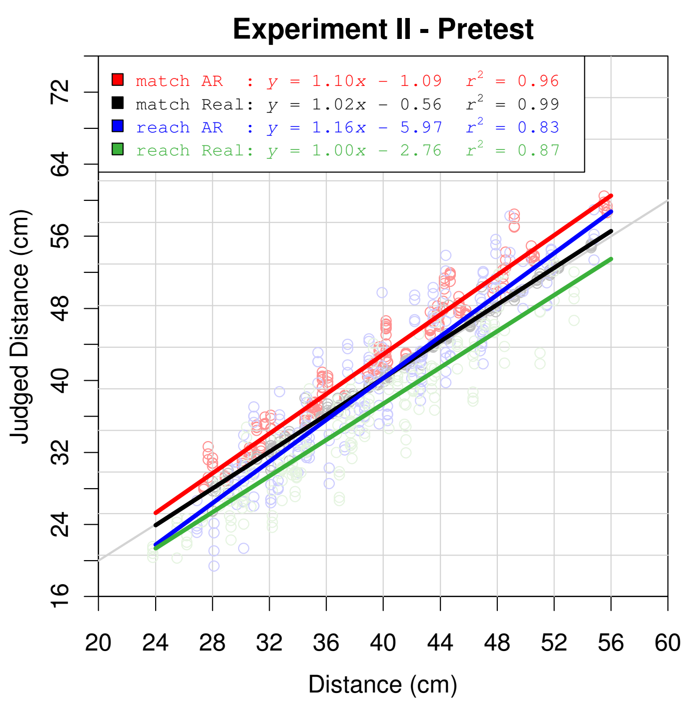


Figure 6.13

Experiment II: Mean judged distance versus the actual distance for the pretest condition.

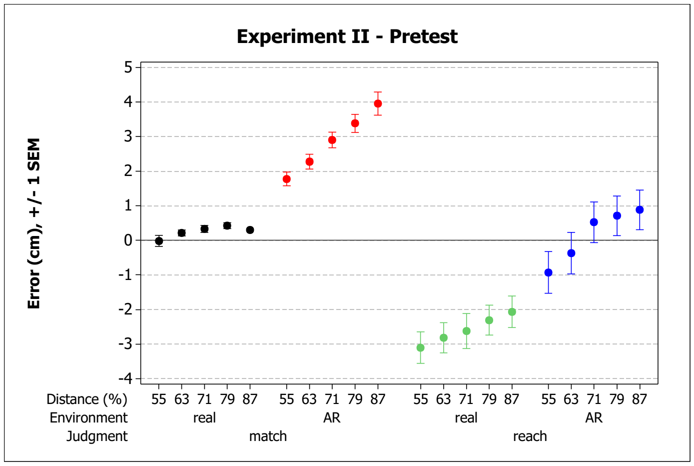


Figure 6.14

Experiment II: Mean error for the depth judgments for the pretest condition.

For (1) the matching, real condition, the judged distance is extremely accurate with a *mean error* of $+0.25$ cm. For (2) the matching, AR condition, the judged distance shows a linear increase with increasing distance, beginning with an *error* of $+1.77$ cm at 55% and progressing to an *error* of $+3.95$ cm at 87% of the maximum reach distance. There is a strong effect of distance on error ($F(4{,}36) = 24.5$, $p < .001$), and a Ryan REGWQ post-hoc homogeneous subset test indicates that every mean is significantly different at $p \leq .05$. The judgment distance means are well-described by a linear model with $m = 1.10$ and $b = -1.09$ ($r^2 = 96\%$); this model indicates that for every additional centimeter of distance, participants localized the target an additional 1 mm farther away than its veridical location. For (3) the reaching, real condition, the judged distance is underestimated, with a mean error of -2.58 cm, with no main effect of distance on error. And, for (4) the reaching, AR condition, the judged distance is slightly overestimated with a mean error of $+0.16$ cm, with no main effect of distance on error. Even though the mean error is really small, the large slope (m =1.16) does not suggest accurate reaching for the AR targets.

6.2.4.2 Overall Results

Figure 6.15 shows the distance judgment error for reach-match-reach (RMR) and match-reach-match (MRM) judgments in the real and the AR environments. Figure 6.16 shows the same results as judged distance versus actual distance. For (1) Real-RMR condition, *match-2* affected the reaching responses; *reach-1* (*mean error* $= -2.58$ cm) and *reach-3* (*mean error* $= -0.77$ cm) are significantly different (F(1,9) = 11.12, $p = .009$). There is no effect of distance on error for any of the *reach-1*, *match-2*, and *reach-3*. For (2) Real-MRM condition, there is no effect of *reach-2* on the matching responses. The distance judgments in *match-1* (mean error $= +0.25$ cm) and *match-3* (mean error $= +0.27$ cm) are not significantly different (F(1,9) = 0.04, $p = .839$).

For (3) AR-RMR condition, there is no effect of *match-2* on reaching judgments; *reach-1* (mean error $= +0.16$ cm) and *reach-3* (mean error $= +0.14$) are not significantly different (F(1,9) = 0.001, $p = .973$). There is a strong main effect of distance on error for *match-2* (F(4,36) = 37.17, $p < .001$), and a Ryan REGWQ post-hoc homogeneous subset test indicates that every mean is significantly different at $p \leq .05$. For (4) AR-MRM condition, there is no effect of *reach-2* on matching judgments; *match-1* and *match-3* are not significantly different (F(1,9) = 0.07, $p = .797$). There are main effect of distance present for both *match-1* (F(4,36) = 24.49, $p < .001$) and *match-3* (F(4,36) = 16.98, $p <$

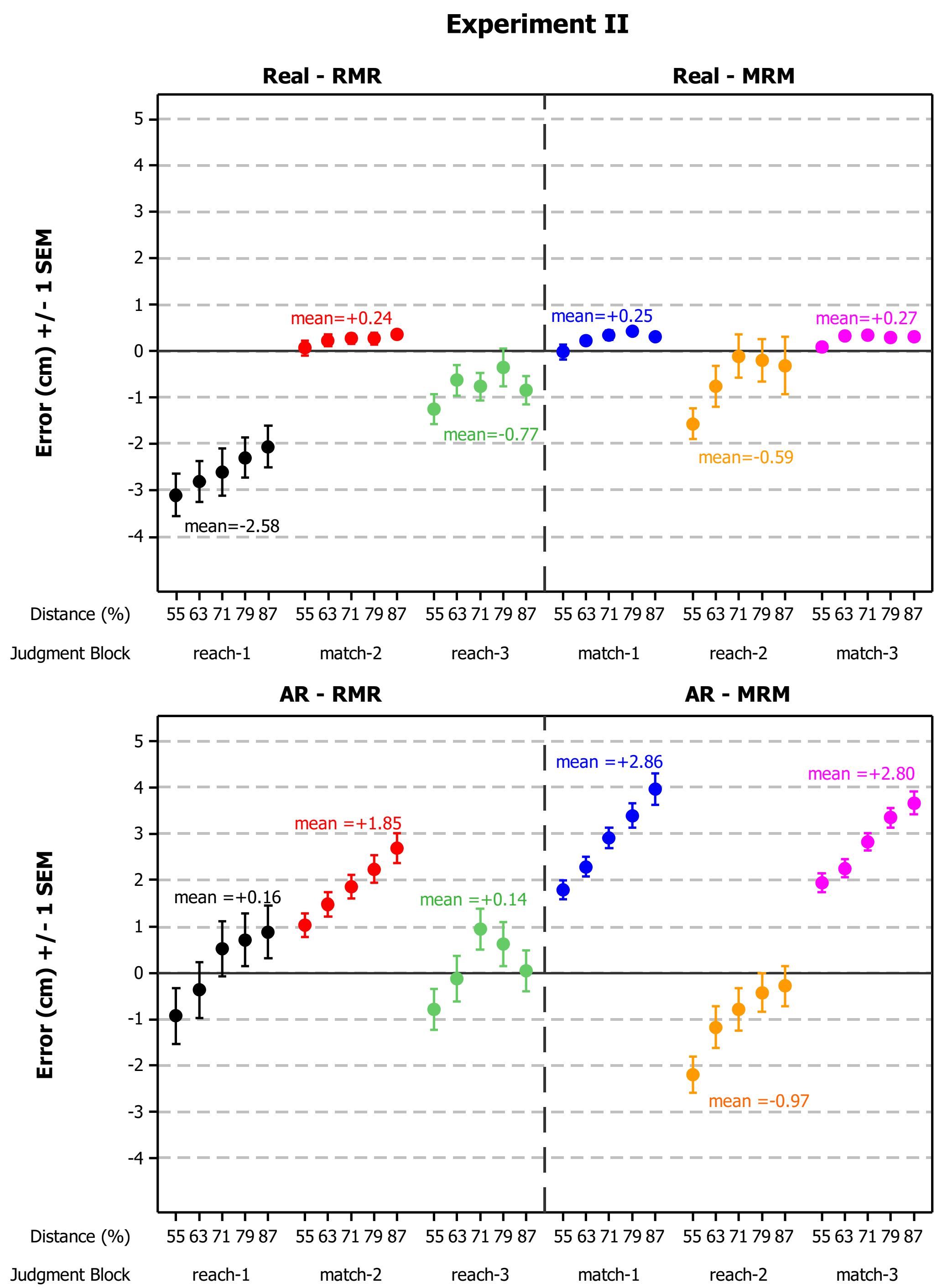


Figure 6.15

Experiment II: Mean error for the depth judgments.

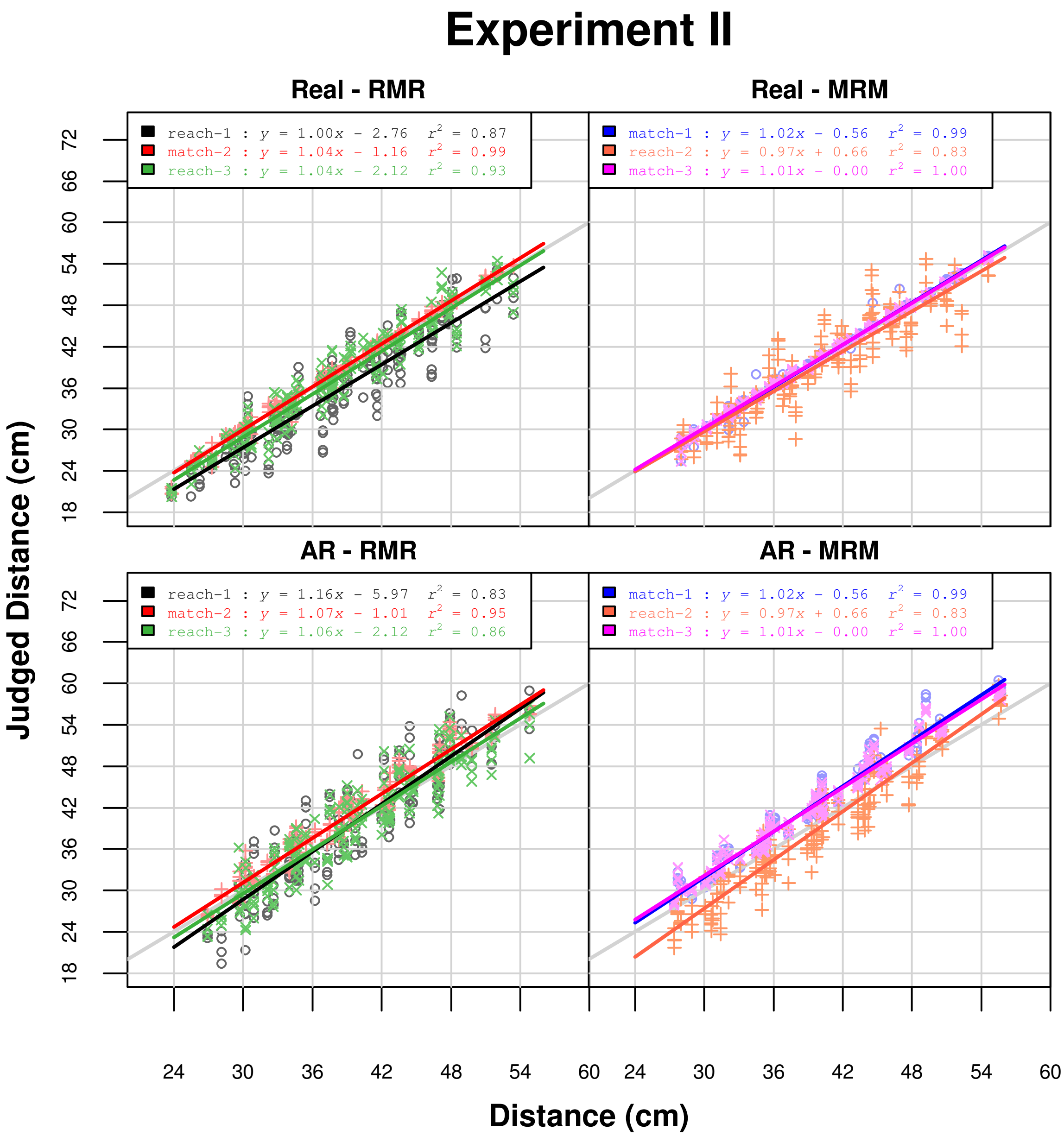


Figure 6.16

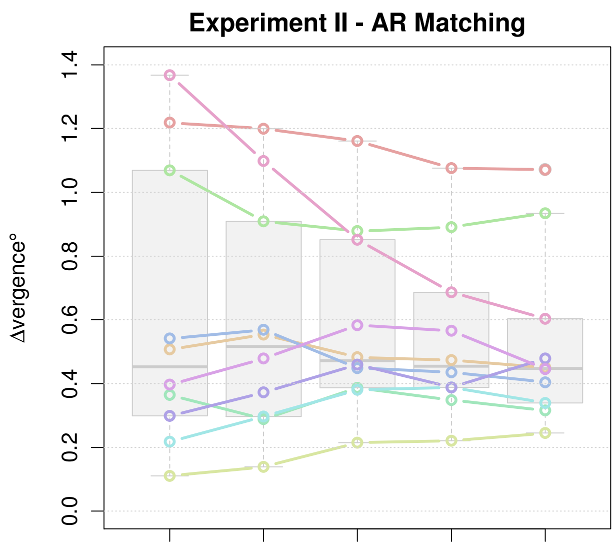
Experiment II: Mean judged distance versus the actual distance.

.001), and Ryan REGWQ post-hoc homogeneous subset tests indicated that every mean is significantly different at $p \leq .05$ for both *match-1*, and *match-3*.

6.2.5 Discussion

The pretest results are directly comparable to Experiment I. In the real environment, the matching responses (*match-1*) were extremely accurate. These results validate the findings of Experiment I, and confirms that participants are very accurate at judging distance to real objects using perceptual matching. Similar to Experiment I, the restricted field-of-view of the HMD did not affect participants' ability to judge distance in the real environment.

In the AR environment, the matching responses were overestimated, and the error increased linearly with increasing distance. As in Experiment I, the data was analyzed to test the hypothesis that the collimated optics of the head-mounted display was driving participant's vergence outwards by a constant angular amount. Figure 6.17 illustrates this hypothesis in terms of the change in vergence angle at each distance for each of the 10 participants in the AR matching condition. For 9 of the 10 participants, ΔV changes less than $0.2°$, and the median line seen in the underprinted boxplot changes less than $0.1°$ with changing distance. The angular changes are small, and indicate that the change in vergence angle is relatively constant with increasing distance. This suggests that the collimated display optics are driving the vergence angle outwards by a constant amount. These results replicate the findings of Experiment I and further emphasize the importance of consistent accommodative and convergence demands for accurate perception of virtual objects at near-field distances.



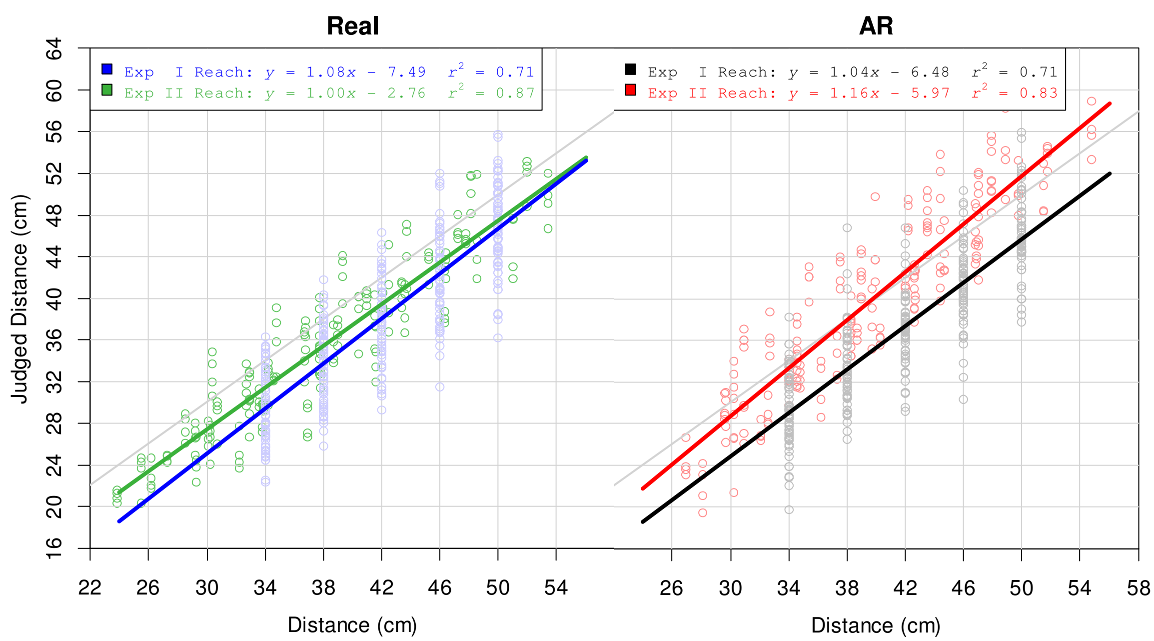


Figure 6.18

Experiment I and Experiment II reaching responses for Real and AR environments.

In Figure 6.18 the green line shows that when blind reaching to real world targets, participants consistently underestimated distances by 2.8 cm. Even though the additional proprioception feedback reduced the underestimation seen in Experiment I (the blue line), these results failed to match the accuracy found by Tresilian and Mon-Williams [87] (Figure 6.8). These results confirm Hypothesis 2.1 that matching is a more accurate distance measure than reaching. Furthermore, these results also confirm Hypothesis 2.2 that the reaching responses in Experiment II are more accurate than Experiment I. In Figure 6.18, the red line shows that when blind reaching AR targets, participants reached $\sim \pm 1$ cm of the actual target distance. While more accurate than the real world results, the large slope (m =1.16) does not suggest accurate reaching for the AR targets.

In the overall analysis, there is an effect of intervention in only one of the four main conditions. In the real, RMR condition, the matching intervention affected the reaching performance. The effect of matching on reaching judgments is also visible in the real, MRM condition, where reach-2 is similar to reach-3 of the real, RMR condition. It is possible that match-1 affected reach-2 in real, MRM condition, in the same way match-2 affected reach-3 in the real, RMR condition. Therefore, the participants were able to correct their reaching responses with the disparity and the proprioception feedback of the matching task. However this effect was not present in the AR condition; there was no significant difference between any of the reach-1, reach-2, and reach-3, indicating perceptual differences in the real and the AR environments. Therefore, Hypothesis 2.3 was rejected as we did not find a consistent effect of the intervention condition. We can sum up the results for the hypotheses that we stated at the beginning:

- **Hypothesis 2.1** was confirmed. The perceptual matching task was more accurate than the blind reaching task for both real and AR environments.
- **Hypothesis 2.2** was confirmed. The reaching responses were more accurate in Experiment II than Experiment I for both real and AR environments.
- **Hypothesis 2.3** was rejected. We did not find a consistent effect of intervention.

6.3 Experiment III - Accommodation

One consistent effect that we found in Experiments I and II was that matching in the AR environment shows a linear increase in the judged distance with increasing target distance, which is consistent with the collimated optics of the head-mounted display driving the participants' vergence angle outwards by a constant angular amount.

Experiment III studied the effect of accommodation on depth perception by comparing three accommodation conditions: *collimated*, *consistent*, and *midpoint*. The collimated condition presented the AR objects at an accommodative demand of infinity, and thus replicated the accommodation settings of Experiments I and II. The consistent condition presented the AR objects at the same accommodative demand as the target distance, and was consistent with how real objects are seen. The midpoint condition presented the AR object with an accommodative demand at the middle of the studied distances. This condition replicated the accommodative settings of current displays that can only utilize a single fixed focal distance. The midpoint condition was included from a practical point of view of display manufacturing: if we find that the distance judgments can be accurately made with an accommodative demand that lies in the middle of the task area, than this finding can be incorporated while designing HMDs.

Experiment III collected data only from younger participants. Because the accommodation ability deteriorates as a person ages, age of the participants is important in studies that examine accommodation. We wanted to study the effect of accommodation in isolation, without any confounding effect of age related changes in accommodation. Therefore, in Experiment III, we only included participants who were less than 40 years old, majority of which ranged in age from 18 to 22. A future experiment will study older participants who wear reading glasses.

Experiement III tested the following hypotheses:

- **Hypothesis 3.1:** ***Overestimation in collimated condition***: Since Experiments I and II found that collimated optics drives participant vergence outwards, we hypothesized the distance judgments would be overestimated in the collimated condition. We also expected to see a similar trend of linear increase in the judged distance as previously seen in Experiments I and II.
- **Hypothesis 3.2:** ***No bias in consistent condition***: Since the consistent condition provides the depth cues that closely replicate real world viewing, we hypothesized that distance judgments would be comparable to real world objects in the consistent condition.
- **Hypothesis 3.3:** ***Bias towards accommodative distance in midpoint condition***: The midpoint condition presented the AR objects with an accommodative demand in the middle of the studied distance range. We hypothesized that the distance judgments will be underestimated for the farther targets, overestimated for the closer targets, and accurate for the middle target.

6.3.1 Experimental Setup

We used the same tabletop apparatus as in Experiments I and II; instead of the head-mounted display, we used an AR-haploscope to present the AR targets. Figure 6.19 shows a side-view diagram of the apparatus and the distance judgment task. Figure 6.20 shows an

Haploscope Experiments III, IV, V

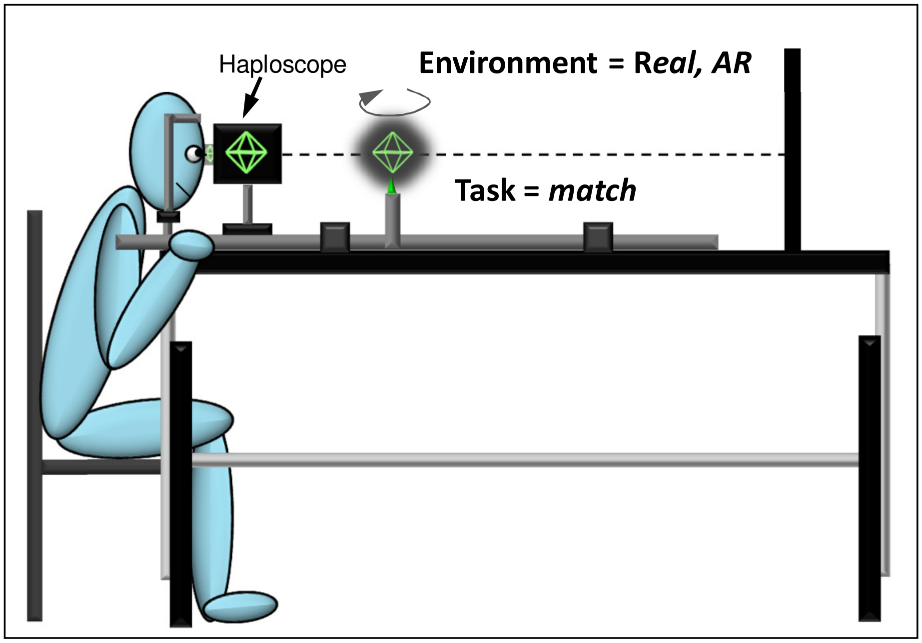


Figure 6.19

Experiment III: Side view diagram of the experimental table.

In the real environment, participants saw a slowly rotating (4 rpm) green wireframe diamond shape with a 5 cm base and 6 cm height. As shown in Figure 6.20, we presented the target object at the eye level; 29 cm above the tabletop. In the AR environment, we presented a virtual rendering of the same target object. In both environments, participants viewed the target object through the haploscope's optics, with no graphics displayed in the real environment. Participants viewed the target object against a black curtain that hung 1.20 m away (see Figure 6.20).

Experiment III, IV, V

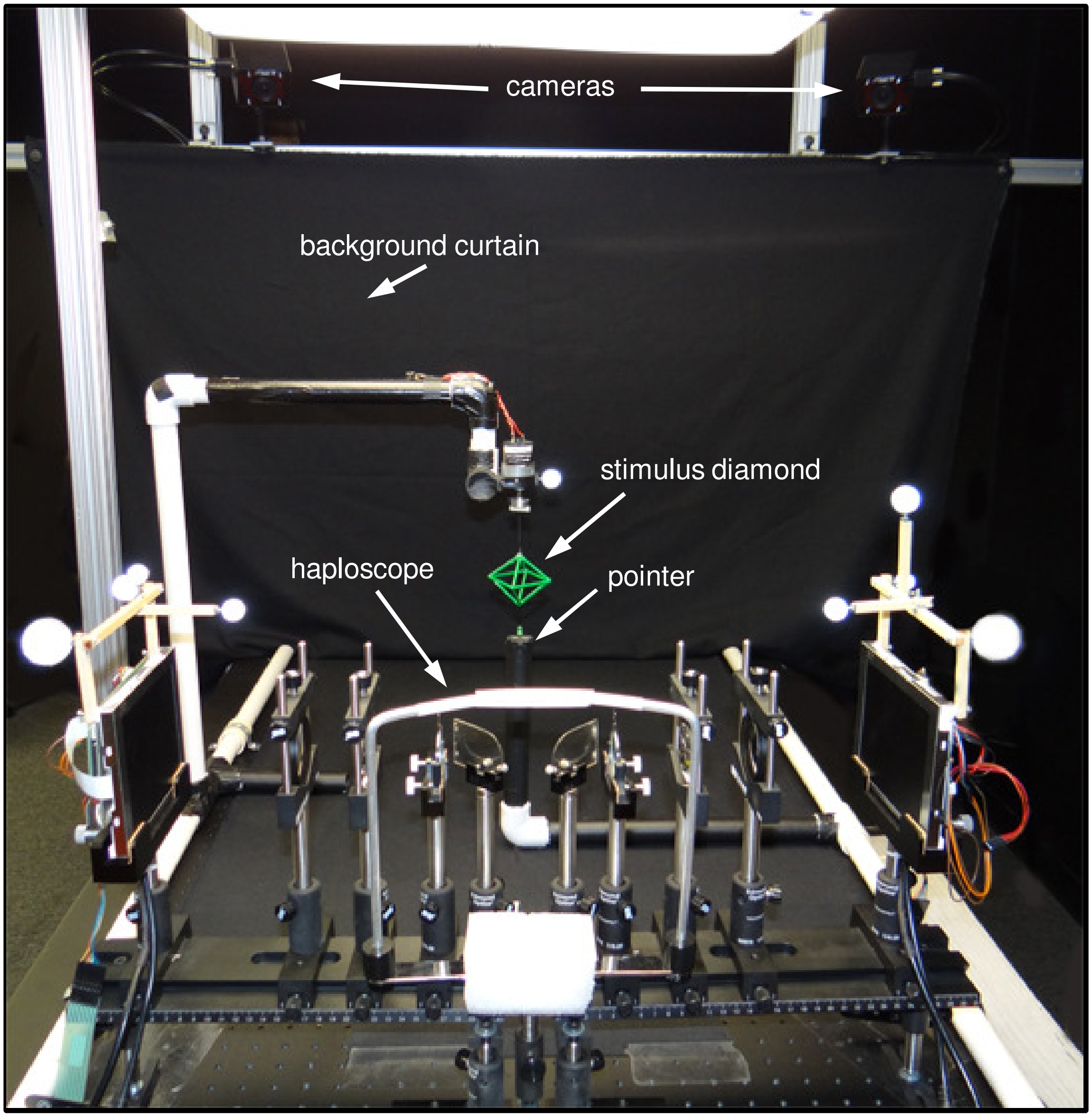


Figure 6.20

Experiment III: Annotated photograph of the experimental table.

We implemented the matching task from Experiment I. We slid a length of white plastic PVC pipe through two collars that were attached to the right-hand side of the table surface (see Figure 6.20). To the pipe, we mounted an arm that extended to the middle of the table. At the end of this arm, we attached a vertical pipe with a green LED on top. While performing the matching task, the participant used their right hand to slide the pipe until the LED was at the same depth as the bottom apex of the rotating target object. We mounted a retro-reflective sphere to the arm, which allowed the tracker to automatically encode the participant's matching responses.

6.3.2 Experimental Variables and Design

Table 6.3 shows the independent and dependent variables of Experiment III. In this study, we recruited 44 younger participants from a population of university students. The participants ranged in age from 18 to 38; the mean age was 20.48, and 25 were female and 19 male. The inter-pupillary distance (IPD) ranged from 54 to 68 mm; the mean IPD was 59.7 mm. 36 participants received course credit and rest received $12 for their participation. We did not analyze data from four participants, as one participant did not pass the stereo vision test and the three did not perform the experiment seriously. As indicated in Table 6.3, we analyzed data from the remaining 40 participants.

Participants performed the perceptual matching task in four accommodation conditions: real, collimated, consistent, and midpoint. As described above, in the real environment, participants saw a physical green octahedron shape as the target object, while in the other three conditions they saw a similar virtual object. We presented the target objects

Table 6.3

Experiment III: Independent and Dependent Variables

INDEPENDENT VARIABLES			
Participant	40		(random variable)
Accommodation	4	between	Real, Collimated, Consistent, Midpoint
Distance	5	within	33.3, 36.4, 40, 44.4, 50 cm
Repetition	6	within	1, 2, 3, 4, 5, 6
DEPENDENT VARIABLES			
Judged Distance			in cm
Error			Judged Distance − Distance

at 5 different distances from the participant: 33.3, 36.4, 40, 44.4, and 50 cm. At 33.3 cm distance the target object subtended a maximum of 10.2° of visual angle, while at 50 cm distance the target object subtended a minimum of 6.8°. Participants saw 6 repetitions of each distance.

As shown in Table 6.3, the primary dependent variable was judged distance, which we measured using the matching task. We also calculated *error* = *judged distance* − *distance*. The sign of the *error* indicated the direction of bias in the distance judgments.

We used a mixed design with the accommodation condition varying between the participants, resulting in four main conditions: (1) real (2) collimated (3) consistent, and (4) midpoint. We varied the presentation order of these four main conditions in a round-robin fashion. For each main condition, our control program generated a list of 5 (*distance*) × 6 (*repetition*) = 30 *distances*. The program then randomly permuted the presentation order of the distances, with the restriction that the same distance could not be presented twice in a row. We collected 1200 data points [40 *participants* × 5 *distances* × 6 *repetitions*].

6.3.3 Procedure

We screened and trained the participants with same procedures as used in Experiment I. We calibrated the haploscope for the participants' IPD. Next, participants wore a pair of plastic safety goggles with circular openings (3.5 cm diameter) that provided a binocular view of the target object through the haploscope's optical combiners, while blocking the view of the rest of the haploscope equipment. Participants looked through the approximate optical centers of each of the optical combiners, while they rested their chin on a chin-rest and placed their forehead against the chin-rest bar (see Figure 6.21).

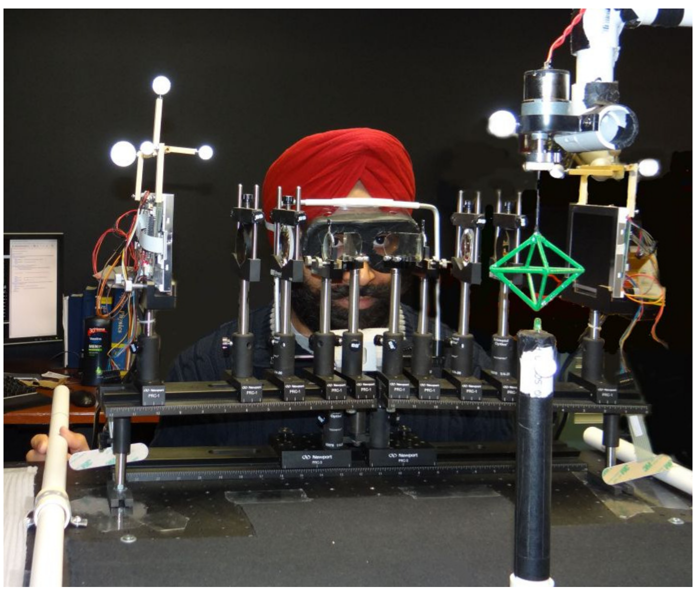


Figure 6.21

Experiment III: A participant after calibration.

The participants next completed one of the four main conditions. At the beginning of the trials, we placed the pointer at a random position between the trackable distance of

23 cm to 67 cm from the observer. For the first trial, the participant adjusted the pointer from this starting position to indicate the distance of the target object. At the beginning of every trial, the participant closed their eyes, and we presented the target object at the trial distance. To present the real target, we adjusted the PVC arm, and to present the AR target, we rotated the left and right image-generator assemblies and switched the appropriate accommodation lenses (see 4.2.2). We followed the same procedure for the real world targets by introducing plain glasses instead of lenses. Next, participants performed the matching task similar to Experiment I.

6.3.4 Results

We analyzed the judged distance results from N = 1200 data values. Figure 6.22 shows the main results as judged distance versus actual distance and Figure 6.23 shows this same data plotted as error. We conducted an ANOVA analysis of this main experimental design. There is a significant 2-way interaction between accommodation and distance (F(12,144) = 5.36, p <.001). This interaction includes a main effect of accommodation on error (F(3,36) = 24.45, p <.001) and a main effect of distance on error (F(4,112) = 3.97, p =.004).

In (1) the real condition, the distance judgments are extremely accurate (*error* = 0.05 *cm*). This result replicated the findings of Experiments I and II that participants are very accurate at judging distances to a real object at near field distances. For (2) the collimated condition, the judged distance shows a linear increase with increasing distance, beginning with an *error* of +0.73 *cm* and progressing to an *error* of +1.78 *cm*. There is also a strong

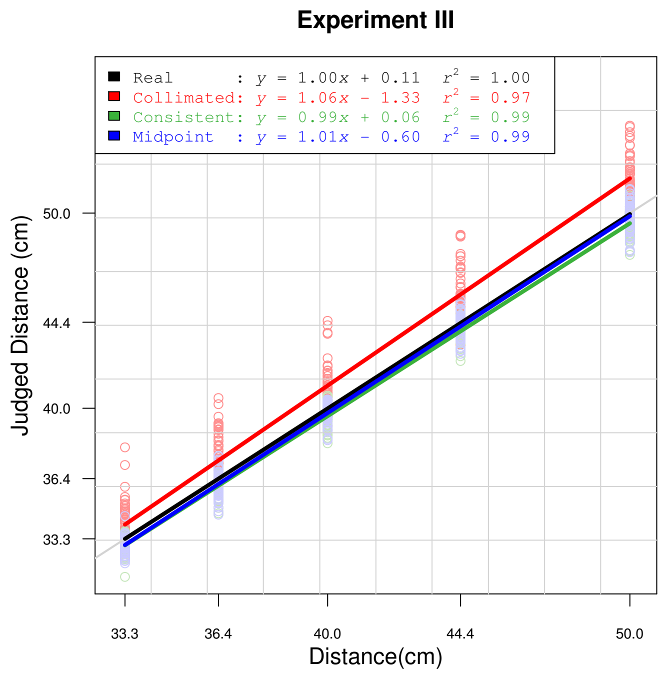


Figure 6.22

Experiment III: Mean judged distance versus the actual distance.

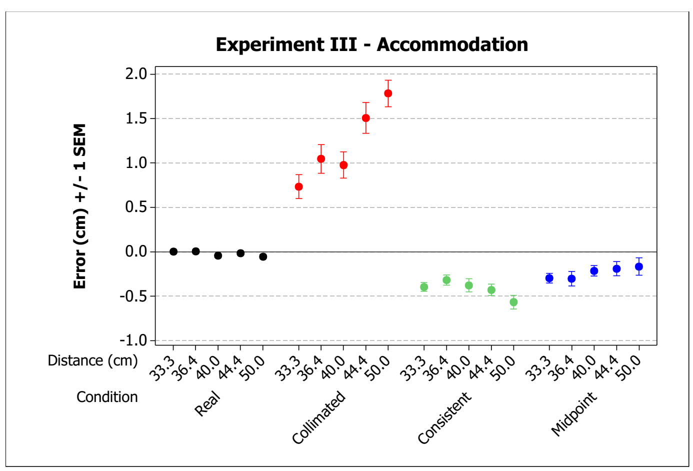


Figure 6.23

Experiment III: Mean error for the depth judgments.

effect of distance on error (F(4,36) = 7.04, $p < .001$), and a Ryan REGWQ post-hoc homogeneous subset test indicates that the response distances fall into two groups: (33.3, 36.4, 40) and (44.4, 50) at $p \leq .05$. The judgment distance means are well-described by a linear model with $m = 1.06$, $b = -1.33$ ($r^2 = 97\%$); this model indicates that for every additional centimeter of distance participants localized the target an additional 0.6 mm farther away than its veridical location. For (3) the consistent condition, the distance judgments are slightly underestimated for all distances (*error* $= -0.42$ *cm*). This condition is significantly different from the real environment (F(1,18) = 10.92, $p = .004$). For (4) the midpoint condition, the distance judgments are also slightly underestimated for all distances with a *mean error* of -0.23 cm (F(1,18) = 1.90, p = .18).

6.3.5 Discussion

In the collimated condition, the distance judgments were overestimated, and the error increased linearly with increasing distance. The disparity analysis of the collimated condition (see Figure 6.24) indicates that the collimated display optics were driving the vergence angle outwards by a constant amount, similar to Experiment I and II. These results confirm Hypothesis 3.1.

In the consistent condition, we expected the distance judgments to be accurate and comparable to the real world. However, the distance judgments were slightly underestimated by 0.42 cm. This underestimation indicates there are other factors besides accommodation and convergence that affect depth perception at near-field distances. We addressed one such factor, brightness, in Experiment IV. These results confirm Hypothesis 3.2 as

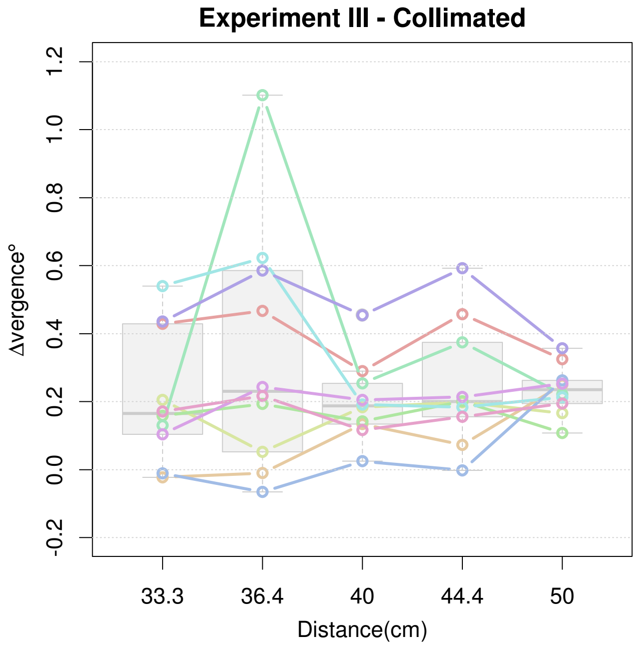


Figure 6.24

Experiment III: The change in vergence angle with distance.

the distance responses were more accurate in the consistent condition as compared to the collimated condition.

In the midpoint condition, the distance responses were extremely accurate. Overall, these results reject Hypothesis 3.3 as we expected that the distances farther than the accommodation distance will be underestimated and the distances closer than the accommodation distance will be overestimated. However, the objects at different distances were not judged differently, and the judgment error did not change with distance. It is possible that the change in accommodation demand was not large enough to affect depth judgments.

An important finding of Experiment III is that there was no significant difference between the distance responses of the consistent and midpoint conditions. These results indicate that even though participants were able to detect the change in accommodation in the midpoint and consistent conditions as compared to the collimated condition, they did not use the differential accommodation information to judge distance at these near-field distances, resulting in similar distance judgment accuracy in the both conditions. This finding is important as it indicates that the human visual system is capable of tolerating the convergence-accommodation mismatch when this mismatch is small. A practical implication of this finding is that single-focal-depth display systems could be used for various near-field AR applications.

These results are consistent with past research on the convergence-accommodation mismatch on depth perception in the near-field (Bingham et al. [10], Rolland et al. [77, 78]). Bingham et al. [10] studied depth perception of virtual objects at 30-40 cm, displayed using a VR display with focal length of 1 meter. They found that accommodation pulled the vergence in towards the virtual image and resulted in distance overestimation. When they approximately matched the accommodative demand (50 cm) with the target distance, the overestimation reduced by 50%. Rolland et al. [77] also found overestimation for AR targets at 80-120 cm using a collimated display. In their next study, the distance judgments became accurate when the targets were presented with correct accommodative demand at 80 cm (Rolland et al. [78]).

Overall, the findings of Experiment III can be summarized as:

- **Hypothesis 3.1** was confirmed. The distance judgments were overestimated in the collimated condition, and error increased linearly with increasing distance, similar to Experiments I and II.
- **Hypothesis 3.2** was confirmed. Making the accommodative demand equal to the target distance resulted in accurate judgments in the consistent condition.

- **Hypothesis 3.3** was rejected, as we did not see the effect of the accommodative demand in the midpoint condition, and the targets at different distances were not judged differently.

6.4 Experiment IV - Brightness

In Experiment III, we expected that the AR objects would be judged accurately in the consistent condition. However, we found underestimation of the AR objects, even when the convergence and accommodative demands were consistent. A potential reason for this effect could be the appearance of the AR objects: the real object was painted with matte green paint, while the AR object was rendered on the monitors, and was brighter than the real object (see Figure 6.25). Since it has been found that brighter objects are perceived as being closer, the difference between the luminance levels of the two objects could be responsible for at least some of the underestimation in perceived distance.

Experiment IV studied the effect of the luminance of the AR object on depth perception. In Experiment III, the luminance of the AR object was 33.19 $cd \cdot m^{-2}$. In Experiment IV, we reduced the luminance of the AR object to 6.84 $cd \cdot m^{-2}$ (see Figure 6.25(c)). The luminance was measured under normal indoor conditions using an ASD FieldSpec Pro spectroradiometer as explained in section 5.2.

Experiment IV had all of the independent and dependent variables as Experiment III. We used the same setup as Experiment III, except for the reduced level of brightness. This experimental design facilitated a direct comparison of the results of Experiment IV and Experiment III. We used the four accommodation conditions of Experiment III as the control conditions in Experiment IV.

nent IV - Brightness

brightness

ghtness (0.4 lux)

ghtness (0.2 lux)

Participants

Matching

tup as Exp III

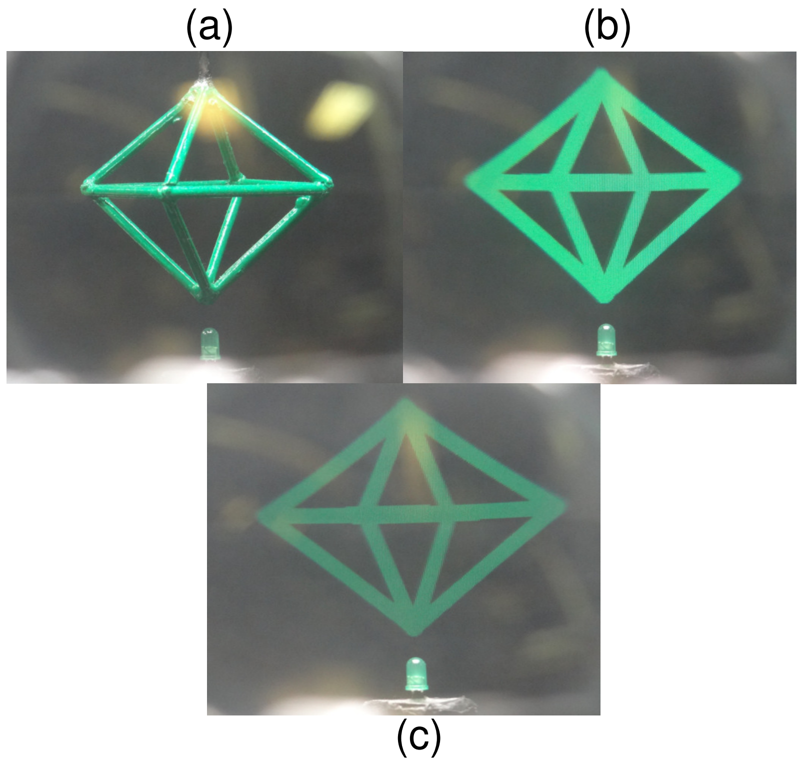


Figure 6.25

Experiment IV: (a) Real world object, (b) bright AR object, and (c) dim AR object.

Experiment IV tested the following hypothesis:

- **Hypothesis 4.1:** ***Bias towards background in all conditions***: We hypothesized that the reduced level of brightness would affect depth perception. We expected that the distance judgments in all of the accommodation conditions will be shifted towards the background, as compared to Experiment III.

6.4.1 Experimental Variables and Design

Table 6.4 shows the independent and dependent variables of Experiment IV. In this study, we recruited 31 younger participants from a population of university students. The participants ranged in age from 17 to 24; the mean age was 19.8, and 10 were female and 21 male. The inter-pupillary distance (IPD) ranged from 52.5 to 67.5 mm; the mean IPD was 60.1 mm. 25 participants received course credit and rest received $12 for their participation. We did not analyze data from one participant, as she did not perform the experiment seriously. We analyzed data from the remaining 30 participants, along with the data from 40 participants from Experiment III.

Table 6.4

Experiment IV: Independent and Dependent Variables

Independent Variables			
Participant	70	random variable (40 from Experiment III)	
Brightness	2	Bright, Dim	
Judgment	4	between	Collimated, Consistent, Midpoint
Distance	5	within	33.3, 36.4, 40, 44.4, 50 cm
Repetition	6	within	1, 2, 3, 4, 5, 6
Dependent Variables			
Judged Distance	in cm		
Error	Judged Distance- Distance		

All the experimental variables were replicated from Experiment III, except for the real-world condition, which was not included as we used Experiment III as a control condition for Experiment IV. We collected 900 data points [*30 participants* $\times$ *5 distances* $\times$ *6 repetitions*].

6.4.2 Procedure

The procedures for this experiment were exactly the same as Experiment III; the same screening and training protocols were used.

6.4.3 Results

We analyzed judged distance results from N = 900 data values. Figure 6.26 shows the main results as judged distance versus actual distance, and Figure 6.27 shows this same data plotted as error. We conducted an ANOVA analysis of this main experimental design. There is a significant 2-way interaction between accommodation and distance (F(8,108) = 3.83, p = .001). This interaction includes a main effect of accommodation on error (F(2,27) =8.35, p = .002), and a main effect of distance on error (F(4,108) = 15.21, $p <$ .001).

For (1) the collimated condition, the judged distance shows a linear increase with increasing distance, beginning with an *error* of $+0.71$ *cm* and progressing to an *error* of $+1.85$ *cm*. There is a strong effect of distance on error (F(4,36) = 11.96, $p <$.001), and a Ryan REGWQ post-hoc homogeneous subset test indicates that the response distances are grouped into three groups: (33.3), (36.4, 40, 44.4), and (50). For (2) the consistent condition, the error in the distance judgments stays constant at 0.1 cm. There is an effect

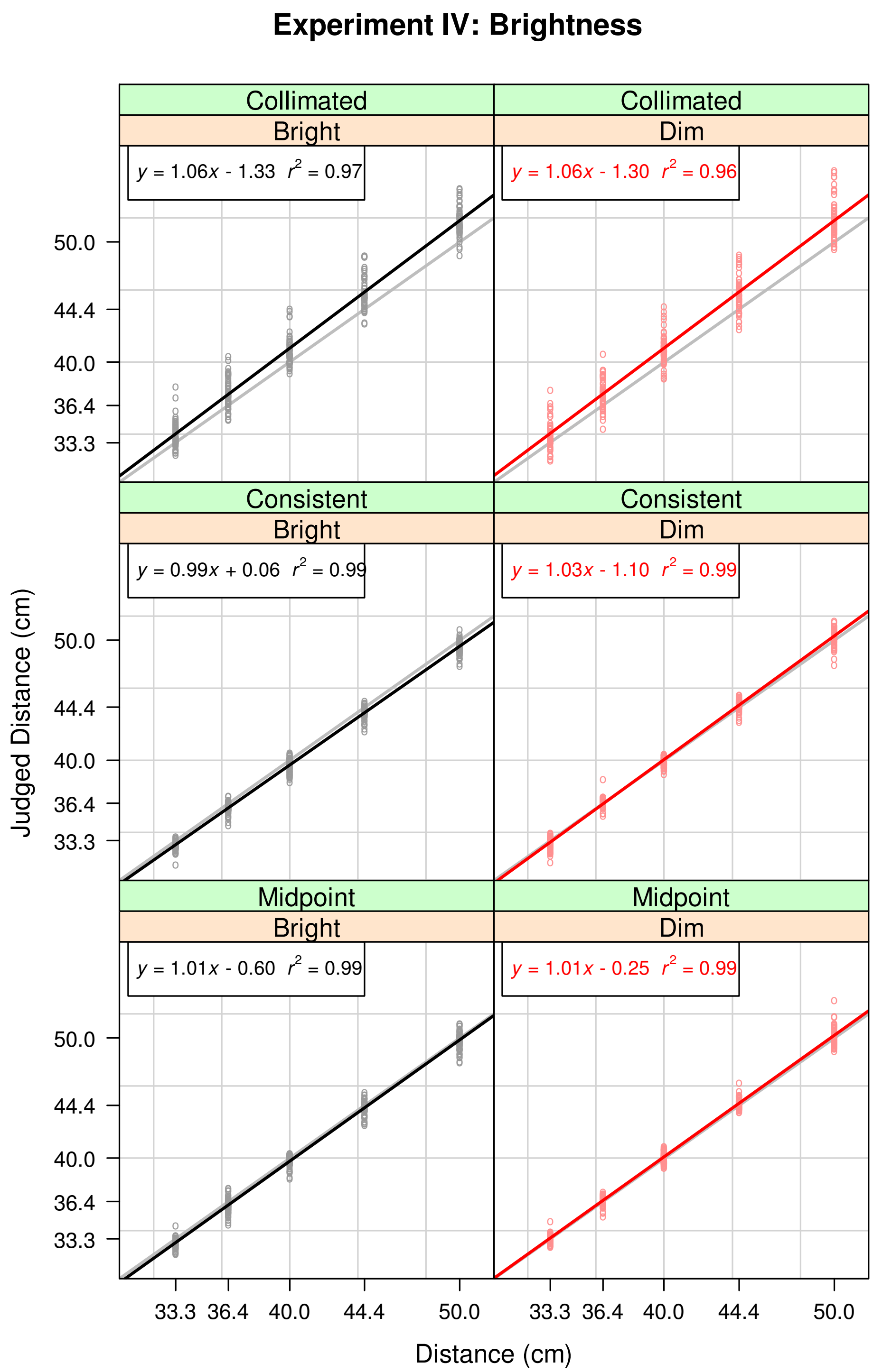


Figure 6.26

Experiment IV: Mean judged distance versus the actual distance.

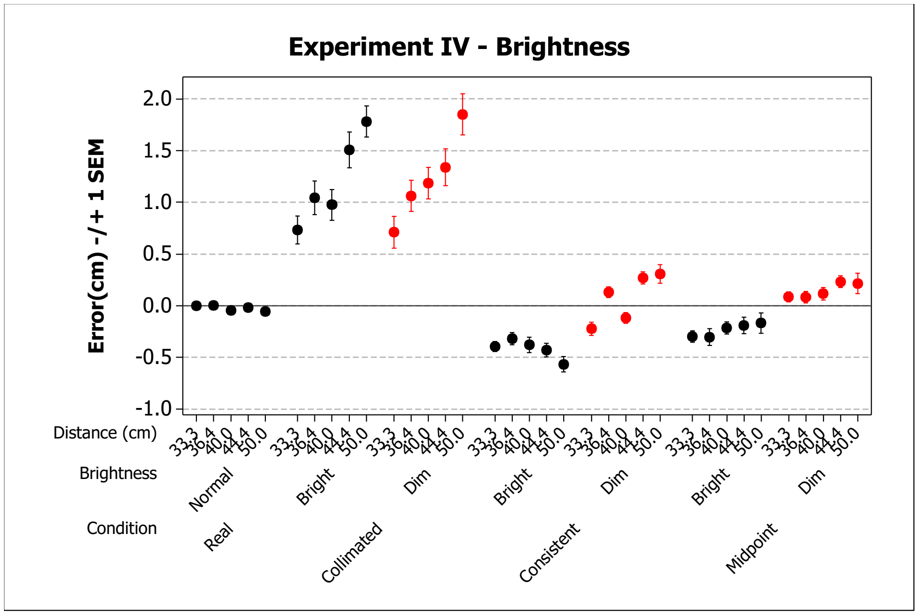


Figure 6.27

Experiment IV: Mean error for the depth judgments for bright and dim targets.

of distance on error (F(4,36) = 6.16, p = .001), and a Ryan REGWQ post-hoc homogeneous subset test indicated that the response distances are grouped into three groups: (33.3, 40.0), (36.4), and (44.4, 50). The target object at 33.3 and 40 cm distance is underestimated, while for other distances it is overestimated. For (3) the midpoint condition, the *error* stays constant at $+0.15$ *cm*, with no effect of distance. The midpoint condition is significantly different from the collimated condition (F(1,18) = 8.42, p =.01) but not from the consistent condition (F(1,18) = 0.25, p =.63).

When comparing the corresponding conditions of Experiment IV and Experiment III, there is no effect of reduced brightness in the collimated condition (F(1,18) = 0.002, p =.096). However, the consistent condition in Experiment IV is significantly different than the consistent condition in Experiment III (F(1,18) = 9.66, p =.006). Also, the midpoint condition in Experiment IV is significantly different than the midpoint condition in Experiment III ($F(1,18) = 4.30,\ p = .053$).

6.4.4 Discussion

In the collimated condition, the distance judgments were overestimated, and the error increased linearly with increasing distance. The disparity analysis (see Figure 6.28) indicates that the collimated display optics were driving the vergence angle outwards by a constant amount; similar to Experiments I, II, and III. However, the distance judgments in this experiment were not significantly different from the distance judgments in the collimated condition of Experiment III.

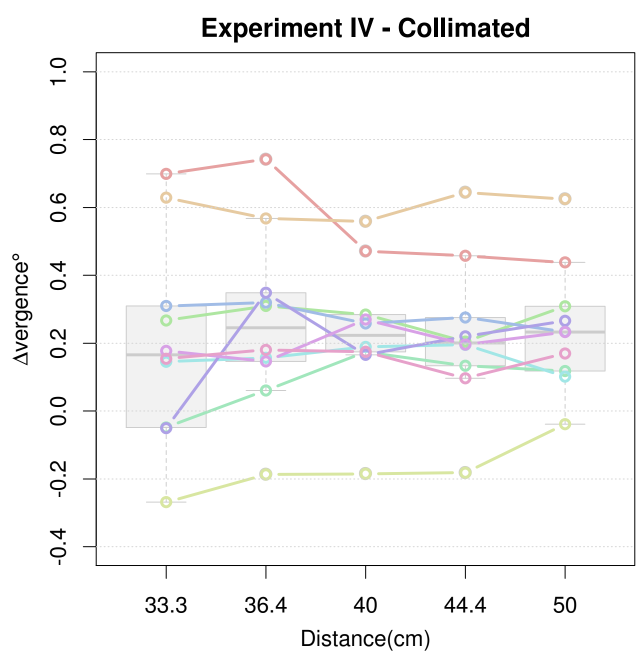


Figure 6.28

Experiment IV: The change in vergence angle with distance.

In the consistent and midpoint conditions, the distance judgments were significantly different from the distance judgments in the corresponding conditions of Experiment III. The distance judgments became more accurate with a mean error of 0.1 cm in the consistent condition and a mean error of 0.15 cm in the midpoint condition. The reduced brightness caused the target object to be perceived farther than in Experiment III.

These results suggest that the brightness affected depth perception in conditions with none (consistent) or very small (midpoint) convergence-accommodation mismatch, and not in the collimated condition with a large convergence-accommodation mismatch. It could mean that the brightness as a depth cue affects depth perception only when it is perceived "correctly" through consistent or nearly consistent vergence and accommodative demands. The human visual system combines the depth information from various depth cues based on their ability to reliably convey depth. It is possible that the level of change in the brightness was only discernible in the consistent and midpoint conditions and this change was not sufficient enough to be detected in the collimated condition. This could lead to the human visual system to not use luminance as a distance cue or at least reduce the weight given to luminance cue during distance estimation.

Our results in the consistent condition replicate the findings of previous research on the effect of brightness in the real environment, with consistent convergence-accommodative demands. Previous studies in real environments have found that increasing the brightness of a target object causes it to be perceived closer for both monocular and binocular conditions (Ashley [4], Farne [27]). To the best of our knowledge, there are no previous studies that tested the effect of brightness for an inconsistent convergence-accommodation pair in the real environment, or that tested the effect of brightness in a virtual or augmented environment.

Overall the findings of Experiment IV can be summarized as:

- **Hypothesis 4.1** was partially confirmed. When compared to Experiment III, the distance judgments were shifted towards the background in the consistent and midpoint conditions, while the distance judgments were not significantly different in the collimated condition.

6.5 Experiment V - Older Participants

Experiment III found an effect of accommodation; participants judged distance accurately when the accommodative demand was equal (consistent) or nearly equal (midpoint) to the stimulus target. In Experiment III, all of the participants were young (mean age of 20.5 years). Because the accommodation ability deteriorates as a person ages, one valid critique of Experiment III is that these results may not generalize to older people.

Our original motivation for studying the depth perception of AR objects is AR-assisted surgery and other AR-enhanced medical applications, which are usually used by surgeons and medical professionals who are in their 40's and older, and who have thus lost substantial accommodative ability relative to people in their early twenties.

Experiment V studied the effect of age on depth perception by studying depth judgments for various accommodation conditions for older participants. We used the same setup as Experiment III. Experiment V had all of the independent and dependent variables as Experiment III. This experimental design facilitated a direct comparison of the results of Experiment V and Experiment III. We used the four accommodation conditions of Experiment III as the control conditions in Experiment V.

Experiment V tested the following hypothesis:

- **Hypothesis 5.1:** ***Older participants less accurate than younger participants*** : Since accommodation ability diminish with age, we hypothesized that the distance judgments for older participants will be less accurate, as compared to the younger participants.

6.5.1 Experimental Variables and Design

Table 6.5 shows the independent and dependent variables of Experiment V. In this study, we recruited 45 older participants from a population of university staff and local community. The participants ranged in age from 41 to 82; the mean age was 56.1, and 23 were female and 22 male. The inter-pupillary distance (IPD) ranged from 53 to 69 mm; the mean IPD was 60.64 mm. All participants received $12 per hour for their participation. 5 participants did not pass stereo vision test. We analyzed data from remaining 40 participants, along with the data from 40 participants from Experiment III.

Table 6.5

Experiment V: Independent and Dependent Variables

INDEPENDENT VARIABLES			
Participant	80	random variable (40 from Experiment III)	
Age	2		Younger, Older
Judgment	3	between	Collimated, Consistent, Midpoint
Distance	5	within	33.3, 36.4, 40, 44.4, 50 cm
Repetition	6	within	1, 2, 3, 4, 5, 6
DEPENDENT VARIABLES			
Judged Distance			in cm
Error			Judged Distance − Actual Distance

All the experimental variables were replicated from Experiment III. We collected 1200 data points [40 (*participant*) $\times$ 5 (*distance*) $\times$ 6 (*repetition*)].

6.5.2 Procedure

The procedures for this experiment were exactly the same as Experiment III; the same screening and training protocols were used.

6.5.3 Results

We analyzed judged distance results from N = 1200 data values. Figure 6.29 and Figure 6.30 show the main results as judged distance versus actual distance. Figure 6.31 shows this same data plotted as error. We conducted an ANOVA analysis of this main experimental design. The main result is a significant 2-way interaction between accommodation and distance (F(12, 144) = 6.10, p <.001).

In (1) the real condition, the distance judgments are extremely accurate (*error* $=$ 0.05 *cm*). This result replicated the findings of Experiments I, II, and III that participants are very accurate at judging distances to a real object at near field distances.

For (2) the collimated condition, the judged distance shows a linear increase with increasing distance, beginning with an $error$ of -0.66 cm and to progressing to an error of +0.74 cm. There is a strong effect of distance on error (F(4, 36) = 6.99, p <.001), and a Ryan REGWQ post-hoc homogeneous subset test indicates that the response distances fall into four groups: (33.3), (36.4), (40, 44.4), and (44.4, 50). The judgment distance means are well described by a linear model with m = 1.08, b = -3.21 (r^2 =97%); this model indicates that for every additional centimeter of distance participants localized the target an

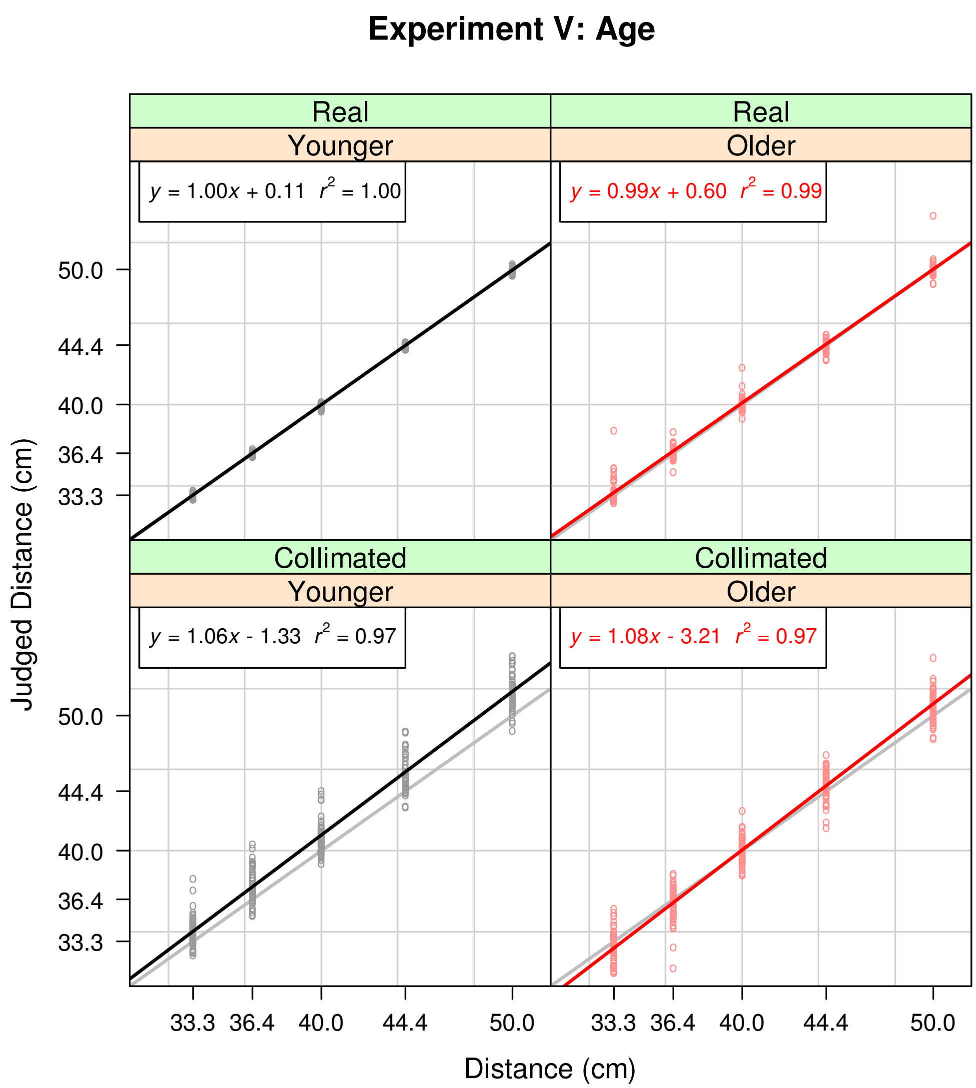


Figure 6.29

Experiment V: Mean judged distance versus the actual distance (Real, Collimated).

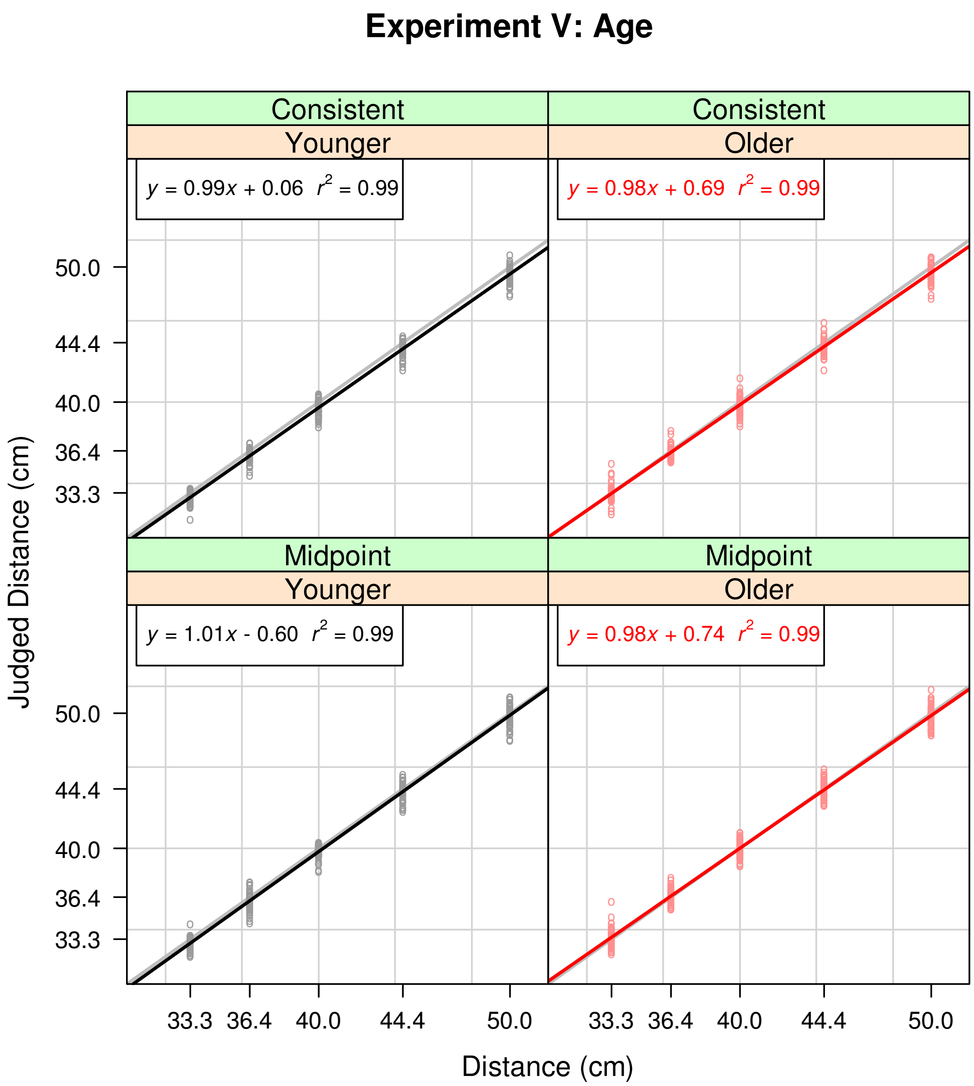


Figure 6.30

Experiment V: Mean judged distance versus the actual distance (Consistent, Midpoint).

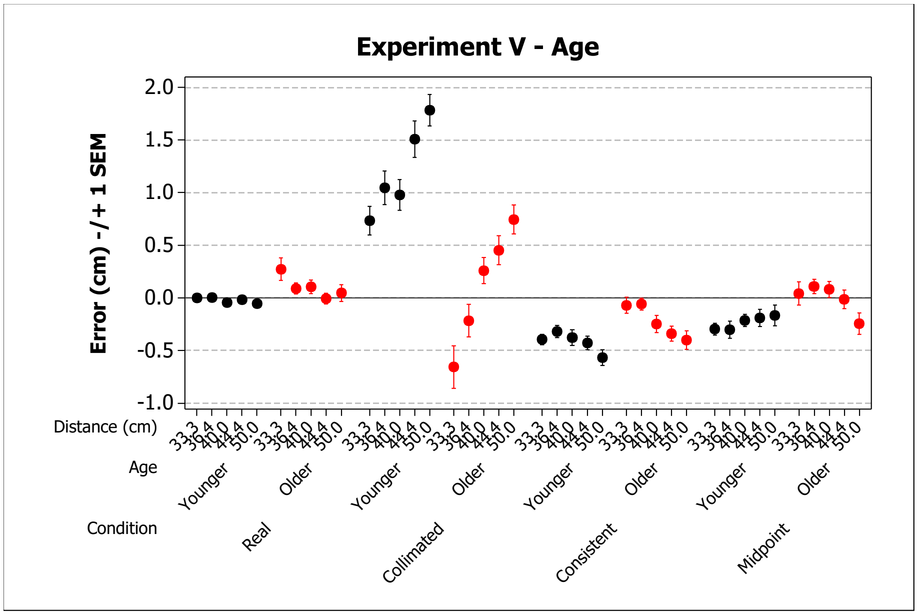


Figure 6.31

Experiment V: Mean error for the depth judgments for younger and older participants.

additional 0.8 mm farther away than its veridical location. For (3) the consistent condition, the distance judgments are slightly underestimated for all distances ($error = -0.22\ cm$). There is no effect of distance for this condition. For (4) the midpoint condition, there is a main effect of distance on error ($F(4, 36) = 2.76,\ p = .043$), and a Ryan REGWQ post-hoc homogeneous subset test indicates that the response distances are grouped into two groups: (33.3, 36.4, 40, 44.4), and (50). The target object at 50 cm distance was underestimated while for other distances it was nearly accurate.

When comparing the corresponding conditions of Experiment V and Experiment III, there is no effect of age in the consistent and midpoint conditions. However the collimated condition in Experiment V is significantly different than the collimated condition in Experiment III ($F(1, 18) = 9.90,\ p = .006$).

6.5.4 Discussion

The results indicated that the older participants judged distance more accurately than younger participants in the collimated condition. There was no difference between the performance of younger and older participants in the consistent and midpoint conditions. These results are surprising in light of many studies that have found decline in visual functions with age (Duane [23], Beers and Heijde [6], Bell et al. [7], Whitbourne [98]). We hypothesized that less sensitivity to accommodation depth cue would result in a decreased performance in depth judgment task for older participants, as the visual problems associated with older age, such as presbyopia, would cause errors in distance judgments.

One potential reason for these results could be a difference in how accommodation is used as a depth cue during distance estimation, by people of different ages. We have consistently found in Experiment I–IV that accommodation affects depth perception indirectly by affecting vergence through the convergence-accommodation reflex (Semmlow and Hung [81], Mon-Williams and Tresilian [62]). Therefore, it is possible that distance judgments of the participants who are more sensitive to accommodation will be affected more by convergence-accommodation reflex, as compared to the participants who are less sensitive to accommodation.

This reasoning could explain the presence of a main effect of accommodation condition for younger participants. Since younger participants are more sensitive to accommodation, the distance responses in the collimated condition were overestimated due to convergence-accommodation mismatch, while distance responses were almost accurate in the consistent and midpoint condition. On the other hand, since older participants are less sensitive to accommodation, there was no main effect of accommodation condition present for older participants. The distance responses in all three accommodation conditions were almost similar.

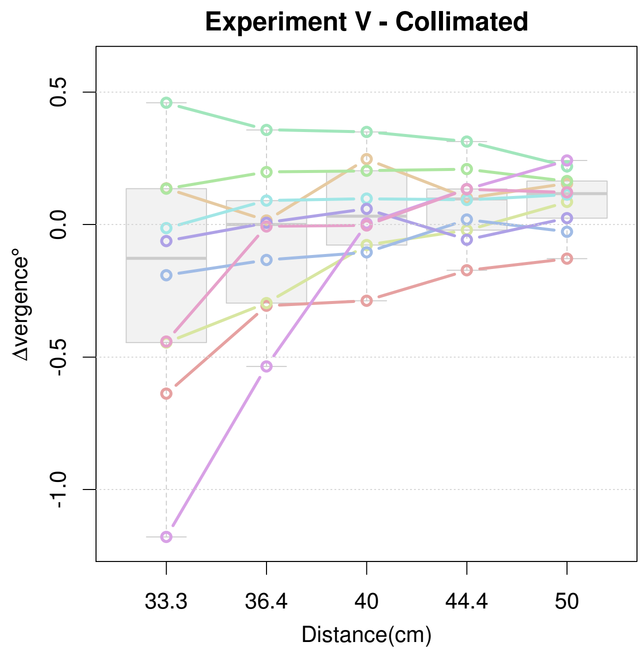


Figure 6.32

Experiment V: The change in vergence angle with distance.

One valid critique of above argument is that there was an effect of distance present in the collimated condition for older participants, indicating at least some sensitivity to accommodation. Similar to Experiment I, II, III, and IV, the error increased linearly with increasing distance in the collimated condition. The disparity analysis (see Figure 6.32) indicates that the collimated display optics were driving the vergence angle outwards by a constant amount.

The presence of a main effect of distance can be explained based on the theory of information integration (see Section 2.2.4), which defines depth cue integration process as a weighted combination of various depth cues. Since older participants are less sensitive to accommodation, less weight was assigned to accommodation depth cue during depth estimation process. The distance judgments depended primarily on the convergence information resulting into no effect of accommodation condition. However, accommodation did contribute in the depth estimation process to some degree, and this contribution was sufficient enough to result in a main effect of distance in the collimated condition.

Our results are consistent with the only experiment that studied the effect of age on depth perception. Bian [8] found that at medium field distances in real world, older participants were relatively more accurate at judging distance than younger participants. They concluded that performance of egocentric distance estimation improves with increased age. To the best of our knowledge, there are no previous studies that tested the effect of age for an augmented reality environment.

Overall the findings of Experiment V can be summarized as:

- **Hypothesis 5.1** was rejected. The older participants judged distance more accurately than the younger participants in the collimated condition. There was no significant difference between the distance judgments of younger and older participants in the consistent and midpoint conditions.

CHAPTER 7

CONCLUSIONS

The author's previous work (Singh et al. [83]) compared two complementary distance judgment protocols of matching and reaching in an AR environment, and found that matching is a more accurate distance measure than reaching. This experiment also found calibration effects over first 5 trials. Two critiques of this work were that (1) this experiment only studied AR targets with no real world ground truth, and (2) this experiment used a within-subjects design, which might have caused the calibration effects.

Experiment I addressed these concerns by replicating the experiment in Singh et al. [83] and compared localization of AR targets with real targets. This experiment also used between-subjects design to alleviate the experimental problems inherently associated with within-subjects design. This experiment confirmed matching to be a more accurate distance measure than reaching. However, this experiment did not find any calibration effect as previously seen in Singh et al. [83]. An important finding of Experiment I was that in the AR condition, the collimated graphics of the display caused an outwards shift in the participants' vergence angle, resulting in a linear increase in error with increasing distance.

In Experiment I, the reaching responses did not show the accuracy seen in previous studies, and were underestimated for both the AR and the real environment. Experiment II tested the hypothesis that these results were due to different bio-mechanical actions between Experiment I and previous studies. Experiment II compared matching and reaching task with a bio-mechanical action similar to Tresilian and Mon-Williams [88], where instead of a physical pointer, participants used the index finger of their right hand to judge distance. This experiment found that the reaching responses in the AR environment became more accurate under the influence of new bio-mechanical action. Furthermore, Experiment II also found that the collimated graphics of the display was causing an outwards shift of the participants' vergence angle in the AR environment, similar to Experiment I.

One consistent effect that was found in Experiments I and II was that matching in the AR environment showed a linear increase in the error with increasing target distance, which is consistent with the infinity accommodation of the collimated optics of the display driving the participants' vergence angle outwards by a constant angular amount. Experiment III studied the effect of accommodation on depth perception by comparing three accommodation conditions: *collimated*, *consistent*, and *midpoint*. Experiment III also found a linear increase in error with increasing distance in the collimated condition, confirming the findings of Experiment I and II. On the other hand, the distance responses in the consistent and midpoint conditions were nearly accurate. These results confirmed that in AR displays, it is important to have consistent or nearly consistent accommodative demand in order to perceive AR objects at correct depth.

In Experiment III, even with consistent accommodative and vergence demands, there was a slight underestimation of depth judgments. Experiment IV tested if the bright appearance of the target object was causing this underestimation, because it has been found that brighter objects are perceived to be closer than their actual position. Experiment IV replicated Experiment III with a relatively dim AR target. Experiment IV found that the distance responses shifted towards background in all conditions. The depth judgments were relatively more accurate in the consistent and midpoint conditions as compared to Experiment III. However, in the collimated condition, the distance judgments were similar to Experiment III and showed similar linear increase in error with increasing distance. These results confirmed that the luminance of target object indeed affects depth judgments, especially when convergence and accommodative demands are equal or nearly equal.

Experiment III found an effect of accommodation, however, all of the participants were young in that experiment. Because the accommodation ability deteriorates as a person ages, the results of Experiment III may not generalize to older people. Experiment V studied the effect of age on depth judgments by replicating Experiment III with older participants. Contrary to the long-held belief, Experiment V found that older participants judged distance more accurately than younger participants.

REFERENCES

[1] G. J. Andersen, A. Saidpour, and M. L. Braunstein, "Effects of collimation on perceived layout in 3-D scenes," *Perception*, vol. 27, 1998, pp. 1305–1315.

[2] N. H. Anderson, "Algebraic Models in Perception," *Handbook of Perception II: Psychophysical Judgment and Measurement*, E. C. Carterette and M. P. Friedman, eds., vol. 2, Academic Press, 1974, pp. 215–298.

[3] N. H. Anderson, *Foundations of Information Integration Theory*, Academic Press, 1981.

[4] M. L. Ashley, "Concerning The Significance of Intensity of Light in Visual Estimates of Depth," *Psychological Review*, vol. 5, no. 6, 1898, pp. 595–615.

[5] R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, 1997, pp. 355–385.

[6] A. P. A. Beers and G. L. Van Der Heijde, "Age Related Changes in The Accommodation Mechanism," *Optometry and Vision Science*, vol. 73, no. 4, 1996, pp. 235–242.

[7] B. Bell, E. Wolf, and C. D. Bernholz, "Depth Perception as a Function of Age," *Aging and Human Development*, vol. 3, 1972, pp. 77–81.

[8] Z. Bian and G. J. Anderson, "Aging and the Perception of Egocentric Distance," *Psychology and Aging*, 2012.

[9] G. P. Bingham, "Calibration of Distance and Size Does Not Calibrate Shape Information: Comparison of Dynamic Monocular and Static and Dynamic Binocular Vision," *Ecological Psychology*, vol. 17, no. 2, Apr 2005, pp. 55–74.

[10] G. P. Bingham, A. Bradley, M. Bailey, and R. Vinner, "Accommodation, Occlusion, and Disparity Matching Are Used to Guide Reaching: A Comparison of Actual Versus Virtual Environments," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 27, no. 6, 2001, pp. 1314–1334.

[11] G. P. Bingham and C. C. Pagano, "The Necessity of a Perception-Action Approach to Definite Distance Perception: Monocular Distance Perception to Guide Reaching," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 24, no. 1, 1998, pp. 145–168.

[12] G. P. Bingham, F. Zaal, D. Robin, and J. A. Shull, "Distortions in Definite Distance and Shape Perception as Measured by Reaching Without and With Haptic Feedback," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 4, 2000, pp. 1436–1460.

[13] E. Brenner and J. M. Van Damme, "Judging Distance from Ocular Convergence," *Vision Research*, vol. 38, no. 4, 1998, pp. 493–498.

[14] B. Bridgeman, S. Peery, and S. Anand, "Interaction of Cognitive and Sensorimotor Maps of Visual Space," *Perception & Psychophysics*, vol. 59, no. 3, Apr 1997, pp. 456–469.

[15] A. Brisbane, "Use a picture, It's worth a thousand words," *Syracuse Post Standard*, March 1911, p. 18.

[16] N. Bruno and J. E. Cutting, "Minimodularity and The Perception of Layout," *Journal of Experimental Psychology: General*, vol. 117, no. 2, Jun 1988, pp. 161–170.

[17] T. P. Caudell and D. W. Mizell, "Augmented Reality: An Application of Heads-Up Display Technology To Manual Manufacturing Processes," *Proceedings: 25th Hawaii International Conference on System Sciences*, 1992, pp. 659–669.

[18] J. Coules, "Effect of Photometric Brightness on Judgments of Distance," *Experimental Psychology*, vol. 50, no. 1, 1955, pp. 19–25.

[19] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson, "The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments," *Perception*, vol. 34, 2005, pp. 191–204.

[20] J. Cutting, "How the eye measures reality and virtual reality," *Behavior Research Methods*, vol. 29, 1997, pp. 27–36.

[21] J. E. Cutting and P. M. Vishton, "Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information About Depth," *Handbook of Perception and Cognition: Perception of Space and Motion*, W. Epstein and S. Rogers, eds., vol. 5, Academic Press, 1995, pp. 69–117.

[22] D. Drascic and P. Milgram, "Perceptual Issues in Augmented Reality," *Stereoscopic Displays and Virtual Reality Systems III*, vol. 2653, 1996, pp. 123–134.

[23] A. Duane, "Normal Values of The Accommodation at All Ages," *Journal of the American Medical Association*, 1912, pp. 1010–1013.

[24] H. Egusa, "Effect of Brightness on Perceived Distance as A Figure-Ground Phenomenon," *Perception*, vol. 11, no. 6, Jan 1982, pp. 671–576.

[25] H. Egusa, "Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual field," *Perception*, vol. 12, no. 2, 1983, pp. 167–175.

[26] S. R. Ellis and B. M. Menges, "Localization of Virtual Objects in the Near Visual Field," *Human Factors*, vol. 40, no. 3, 1998, pp. 415–431.

[27] M. Farne, "Brightness as An Indicator to Distance: Relative Brightness Per se or Contrast with The Background?," *Perception*, vol. 6, 1977, pp. 287–293.

[28] E. F. Fincham and J. Walton, "The Reciprocal Actions of Accommodation and Convergence," *Journal of Physiology*, vol. 137, no. 1957, 1957, pp. 488–508.

[29] S. K. Fisher and K. J. Ciuffreda, "Accommodation and Apparent Distance," *Perception*, vol. 5, 1988, pp. 609–621.

[30] J. M. Foley, "Effect of distance information and range on two indices of visually perceived distance," *Perception*, vol. 6, 1977, pp. 449–460.

[31] J. M. Foley, "Binocular Distance Perception," *Psychological Review*, vol. 87, no. 5, 1980, pp. 411–434.

[32] J. M. Foley and R. Held, "Visually Directed Pointing as A Function of Target Distance, Direction, and Available Cues," *Perception & Psychophysics*, vol. 12, no. 3, 1972, pp. 263–268.

[33] G. A. Fry, "Further Experiments on The Accommodation-Convergence Relationship," *American Journal of Optometry*, vol. 16, 1939, pp. 325–336.

[34] B. Gillam, "The Perception of Space Layout," *Perception of Space and Motion*, W. Epstein and S. Rogers, eds., 2 edition, Academic Press, 1995, pp. 23–67.

[35] W. C. Gogel, "The Common Occurrence of Errors of Perceived Distance," *Perception & Psychophysics*, vol. 25, no. 1, Jan 1979, pp. 2–11.

[36] W. C. Gogel and J. D. Tietz, "A Comparison of Oculomotor and Motion Parallax Cues of Egocentric Distance," *Vision Research*, vol. 19, 1979, pp. 1161–1170.

[37] M. A. Goodale, "Transforming Vision into Action," *Vision Research*, vol. 51, no. 13, Jul 2011, pp. 1567–1587.

[38] M. A. Goodale and A. D. Milner, "Separate Visual Pathways for Perception and Action," *Trends in Neurosciences*, vol. 15, no. 1, Jan 1992, pp. 20–25.

[39] C. S. Harris, "Perceptual Adaptation To Inverted, Reversed, and Displaced Vision," *Psychological Review*, vol. 72, no. 6, Nov 1965, pp. 419–44.

[40] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue," *Journal of Vision*, vol. 8(3), no. 33, 2008, pp. 1–30.

[41] V. Interrante, L. Anderson, and B. Ries, "Distance Perception in Immersive Virtual Environments , Revisited," *Proceedings: IEEE conference on Virtual Reality (VR 06)*, 2006, pp. 3–11.

[42] J. A. Jones, J. E. Swan II, G. Singh, E. Kolstad, and S. R. Ellis, "The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception," *Proceedings: ACM Symposium on Applied Perception in Graphics and Visualization (APGV 2008)*, Los Angeles, CA, USA, 2008, pp. 9–14.

[43] S. Kasthurirangan and A. Glasser, "Age Related Changes in Accommodative Dynamics in Humans," *Vision research*, vol. 46, Apr 2006, pp. 1507–1519.

[44] R. V. Kenyon, K. J. Ciuffreda, and L. Stark, "Binocular Eye Movements During Accommodative Vergence," *Vision Research*, vol. 18, 1978, pp. 545–555.

[45] D. Kersten and G. E. Legge, "Convergence Accommodation," *Journal of the Optical Society of America*, vol. 73, no. 3, Mar 1983, pp. 332–338.

[46] J. M. Knapp, *The Visual Perception of Egocentric Distance in Virtual Environments*, doctoral dissertation, University of California, Santa Barbara, Santa Barbara, California, 1999.

[47] J. M. Knapp and J. M. Loomis, "Limited Field of View of Head-Mounted Displays Is Not the Cause of Distance Underestimation in Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 13, no. 5, 2004, pp. 572–577.

[48] T. Künnapas, "Distance Perception as A Function of Available Visual Cues," *Journal of Experimental Psychology*, vol. 77, no. 4, Aug 1968, pp. 523–529.

[49] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and Modeling of Depth Cue Combination: In Defense of Weak Fusion," *Vision research*, vol. 35, no. 3, Feb 1995, pp. 389–412.

[50] W. S. Levine, *The Control Handbook*, 1 edition, CRC-Press, New York, 1996.

[51] J. M. Loomis and J. M. Knapp, "Visual Perception of Egocentric Distance in Real and Virtual Environments," *Virtual and Adaptive Environments: Applications, Implications, and Human Performance*, L. J. Hettinger and M. Haas, eds., CRC Press, Mahwah, NJ, 2003, chapter 2, pp. 21–46.

[52] E. E. Maddox, “Investigations on The Relationship Between Convergence and Accommodation of The Eyes,” *Journal of Anatomy and Physiology*, vol. 20, 1886, pp. 475–508, 565–584.

[53] D. W. Massaro, “Ambiguity in Perception and Experimentation,” *Journal of Experimental Psychology: General*, vol. 117, no. 4, 1988, pp. 417–421.

[54] L. I. N. Mazyn, M. Lenoir, G. Montagne, C. Delaey, and G. J. P. Savelsbergh, “Stereo Vision Enhances The Learning of A Catching Skill,” *Experimental Brain Research*, vol. 179, no. 4, Jun 2007, pp. 723–726.

[55] J. W. McCandless, S. R. Ellis, and B. D. Adelstein, “The Effect of Accommodative Demand, Motion Parallax and Age on Virtual Object Localizations,” *Proceedings: 43th Annual Meeting of Human Factors and Ergonomics Society*, 1999, pp. 1328–1332.

[56] J. W. McCandless, S. R. Ellis, and B. D. Adelstein, “Localization of a Time-Delayed, Monocular Virtual Object Superimposed on a Real Environment,” *Presence*, vol. 9, no. 1, 2000, pp. 15–24.

[57] E. McCurdy, *The Notebooks of Leonardo da Vinci, Volume II*, Reynal & Hitchcock, 1938.

[58] P. Milgram and F. Kishino, “A Taxonomy of Mixed Reality Visual Displays,” *IEICE Transactions on Information Systems*, vol. E77-D, no. 12, 1994, pp. 1312–1329.

[59] M. Mon-Williams and G. P. Bingham, “Calibrating Reach Distance to Visual Targets,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 33, no. 3, 2007, pp. 645– 656.

[60] M. Mon-Williams and J. R. Tresilian, “The Size-Distance Paradox is A Cognitive Phenomenon,” *Experimental Brain Research*, vol. 126, no. 4, 1999, pp. 578–582.

[61] M. Mon-Williams and J. R. Tresilian, “Some recent studies on the extraretinal contribution to distance perception,” *Perception*, 1999, pp. 167–181.

[62] M. Mon-williams and J. R. Tresilian, “Ordinal Depth Information From Accommodation?,” *Ergonomics*, vol. 43, no. 3, 2000, pp. 391–404.

[63] M. Mon-Williams, J. R. Tresilian, R. D. Mcintosh, and A. D. Milner, “Monocular and Binocular Distance Cues : Insights From Visual Form Agnosia I (of III),” *Experimental Brain Research*, vol. 139, 2001, pp. 127–136.

[64] M. Mon-Williams, J. P. Wann, M. Jenkinson, and K. Rushton, “Synaesthesia in The Normal Limb,” *Proceedings of the Royal Society, Series B*, vol. 264, Jul 1997, pp. 1007–1010.

[65] J. Muller, *Elements of Physiology. Vol. II*, Taylor and Walton. Translated in 1842 by William Baly, London, 1826.

[66] A. Naceri, R. Chellali, and T. Hoinville, "Depth Perception within Peripersonal Space using Head Mounted Display," *Presence: Teleoperators and Virtual Environments*, vol. 20, no. 3, 2011, pp. 254–272.

[67] S. Nagata, "How to Reinforce Perception of Depth in Single Two-Dimensional Pictures," *Pictorial Communication in Virtual and Real Environments*, S. R. Ellis, M. Kaiser, and A. J. Grunwald, eds., 2 edition, London: Taylor and Francis, 1993, pp. 527–545.

[68] P. E. Napieralski, B. M. Altenhoff, J. W. Bertrand, L. O. Long, S. V. Babu, C. C. Pagano, J. Kern, and T. A. Davis, "Near-Field Distance Perception in Real and Virtual Environments Using Both Verbal and Action Responses," *ACM Transactions on Applied Perception*, vol. 8, no. 3, Aug 2011, pp. 1–19.

[69] D. Owens and H. Leibowitz, "Oculomotor adjustments in darkness and the specific distance tendency," *Perception & Psychophysics*, vol. 20, 1976, pp. 2–9.

[70] D. A. Owens and H. W. Liebowitz, "Accommodation, Convergence, and Distance Perception in Low Illumination," *American Journal of Optometry & Physiological Optics*, vol. 57, no. 9, 1980, pp. 540–550.

[71] C. C. Pagano and G. P. Bingham, "Comparing Measures of Monocular Distance Perception: Verbal and Reaching Errors are Not Correlated," *Journal of Experimental Psychology. Human Perception and Performance*, vol. 24, no. 4, 1998, pp. 1037–1051.

[72] J. Paillard and M. Brouchon, "Active and Passive Movements in The Calibration of Position Sense," *The Neuropsychology of Spatially Oriented Behavior*, Dorsey Press, 1968, chapter 3, pp. 37–55.

[73] E. Peli, "Optometric and Perceptual Issues with Head-Mounted Displays," *Visual Instrumentation: Optical Design and Engineering Principles*, P. Mouroulis, ed., McGraw-Hill, 1999, pp. 205–276.

[74] C. Prablanc, J. F. Echallier, E. Komilis, and M. Jeannerod, "Optimal Response of Eye and Hand Motor Systems in Pointing at a Visual Target: I. Spatio-Temporal Characteristics of Eye and Hand Movements and Their Relationships when Varying the Amount of Visual Information," *Biological Cybernetics*, vol. 35, no. 2, 1979, pp. 113–124.

[75] F. H. Previc, "The Neuropsychology of 3-D Space," *Psychological Bulletin*, vol. 124, no. 2, Sep 1998, pp. 123–164.

[76] H. Ripps, N. B. Chin, I. M. Siegel, and G. M. Breinin, "The Effect of Pupil Size on Accommodation, Convergence, and The AC/A Ratio.," *Investigative Ophthalmology*, vol. 1, Feb 1962, pp. 127–135.

[77] J. P. Rolland, W. Gibson, and D. Ariely, "Towards Quantifying Depth and Size Perception in Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 4, no. 3, 1995, pp. 24–49.

[78] J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci, "Method of Adjustment versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Helmet-Mounted Displays," *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 6, 2002, pp. 610–625.

[79] A. D. Ryer, *Light Measurement Handbook*, 1st edition, International Light Inc., 1998.

[80] E. Schmidt, "Every 2 Days We Create As Much Information As We Did Up To 2003," 2010, http://techcrunch.com/2010/08/04/schmidt-data/.

[81] J. L. Semmlow and G. K. Hung, "The Near Response : Theories of Control," *Vergence Eye Movements: Basic & Clinical Aspects*, C. M. Schor and K. J. Ciuffreda, eds., Butterworth, 1983, pp. 175–195.

[82] G. Singh, *Near-Field Depth Perception in See-Through Augmented Reality*, master's thesis, Mississippi State University, Mississippi State, Mississippi, 2010.

[83] G. Singh, J. E. Swan II, J. A. Jones, and S. R. Ellis, "Depth Judgment Measures and Occluding Surfaces in Near-Field Augmented Reality," *Applied Perception in Graphics and Visualization (APGV 2010)*, Los Angeles, California, USA, 2010, pp. 149–156.

[84] J. E. Swan II, J. A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, "Egocentric Depth Judgments in Optical, See-Through Augmented Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 3, 2007, pp. 429–442.

[85] J. E. Swan II, M. A. Livingston, H. S. Smallman, D. Brown, J. L. Gabbard, and D. Hix, "A Perceptual Matching Technique for Depth Judgments in Optical, See-Through Augmented Reality," *Proceedings: The IEEE conference on Virtual Reality (VR 06)*, 2006, pp. 19–26.

[86] I. L. Taylor and F. C. Sumner, "Actual Brightness and Distance of Individual Colors When Their Apparent Distance is Held Constant," *Journal of Psychology*, vol. 19, 1945, pp. 79–85.

[87] J. R. Tresilian and M. Mon-williams, "A Curious Illusion Suggests Complex Cue Interactions in Distance Perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, no. 3, 1999, pp. 677–687.

[88] J. R. Tresilian, M. Mon-Williams, and B. M. Kelly, "Increasing Confidence in Vergence as A Cue to Distance," *Royal Society of London*, 1999, pp. 39–44.

[89] M. T. Turvey and J. Solomon, "Visually Perceiving Distance: A Comment on Shebilske, Karmiohl, and Proffitt (1983)," *Journal of Experimental Psychology. Human Perception and Performance*, vol. 10, no. 3, Jun 1984, pp. 449–454.

[90] R. J. van Beers, A. C. Sittig, and J. J. Denier van der Gon, "How Humans Combine Simultaneous Proprioceptive and Visual Position Information," *Experimental Brain Research*, vol. 111, no. 2, Sep 1996, pp. 253–61.

[91] A. Viguier, G. Clment, and Y. Trotter, "Distance Perception Within Near Visual Space," *Perception*, vol. 30, 2001, pp. 115–124.

[92] C. von Hofsten and B. Rosblad, "The Integration of Sensory Information in The Development of Precise Manual Pointing," *Neuropsychologia*, vol. 26, 1988, pp. 805–821.

[93] H. Wallach and L. Floor, "The Use of Size Matching to Demonstrate The Effectiveness of Accommodation and Convergence as Cues for Distance," *Perception & Psychophysics*, vol. 10, no. 6, 1971, pp. 423–428.

[94] J. P. Wann, "The Integrity of Visual-Proprioceptive Mapping in Cerebral Palsy," *Neuropsychologia*, vol. 29, 1991, pp. 1095–1106.

[95] J. P. Wann and S. F. Ibrahim, "Does Limb Proprioception Drift?," *Experimental Brain Research*, vol. 91, no. 1, Jan 1992, pp. 162–166.

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

[97] C. Wheatstone, "On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision," *Philosophical Transactions of The Royal Society*, vol. 33, 1838, pp. 371–394.

[98] S. K. Whitbourne, *The Aging Individual: Physical and Psychological Perspectives*, 2nd edition, Springer, 2002.

[99] 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," *Applied Perception in Graphics and Visualization*, 2004, pp. 35–38.

[100] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson, "The Effects of Head-Mounted Display Mechanical Properties and Field of View on Distance Judgments in Virtual Environments," *ACM Transactions on Applied Perception*, vol. 6, March 2009, pp. 8:1–8:14.

[101] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-regehr, "Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 17, no. 1, 2008, pp. 91–101.

[102] R. S. Woodworth, *The Accuracy of Voluntary Movement*, doctoral dissertation, Columbia University, New York, Jul 1899.

[103] B. Wu, T. L. Ooi, and Z. J. He, "Perceiving Distance Accurately By a Directional Process of Integrating Ground Information," *Nature*, vol. 428, no. March, 2004, pp. 73–78.