THESIS

EXAMINING EYE VERGENCE ANGLE IN OPTICAL SEE-THROUGH AUGMENTED REALITY

Submitted by
Matthew Sturgeon
Department of Computer Science

In partial fulfillment of the requirements
For the Degree of Master of Science
Colorado State University
Fort Collins, Colorado
Fall 2025

Master's Committee:

Advisor: Mohammed Safayet Arefin

Mohammed Safayet Arefin
Jerald L. Thomas
Anne Cleary

Copyright by Matthew Sturgeon 2025

All Rights Reserved

ABSTRACT

EXAMINING EYE VERGENCE ANGLE IN OPTICAL SEE-THROUGH AUGMENTED REALITY

A primary objective of optical see-through (OST) Augmented Reality (AR) systems is to enable users to perceive the depth of AR objects with the same accuracy as they perceive the depth of real-world objects. Previous studies have shown that individuals often make depth judgments that either underestimate or overestimate the depth of AR objects compared to real-world objects. Recently, OST AR devices have incorporated eye-tracking technology, offering the opportunity to objectively measure and investigate the depth-dependent components of the human visual system (e.g., eye vergence angle) while perceiving the depth of real and AR objects. This paper measures and examines the eye vergence angle (EVA) for both real and AR objects at four different depths (0.35 m, 0.75 m, 1.5 m, and 4.0 m) using integrated eye-tracking systems of the two commercial devices: Microsoft HoloLens 2 and Magic Leap 2. The experiment considered a four-alternative forced-choice visual discrimination task with a repeated-measures design, involving 24 participants. Our findings showed that subjective (verbal estimation) and objective (EVA) measures of depth perception were consistent with the depth of both real and AR objects in both OST AR devices. The results demonstrated individual differences in EVA for each device across various depths. No difference in EVA was identified between real and AR objects. Additionally, the EVA range differs between the OST AR devices for each depth. These findings indicate that objective EVA measurements can not be generalized across different OST AR devices and individuals.

ACKNOWLEDGEMENTS

I first would like to thank all of my coauthors who made our published work possible "Examining Eye Vergence During Perceived Depth Changes with Eye Tracking System in Optical See-Through Augmented Reality" which makes up a majority of the content in this thesis, Dakota-Kenoyer Healy, Russel Cohen Hoffing, Steven Thurman and Dr Mohammed Safayet Arefin. All of their hard work and dedication throughout the process has allowed this work to be what it is today.

Many thanks to Dr Mohammed Safayet Arefin for all of his enthusiastic support and guidance that pushed me to achieve so much throughout my Masters. I feel that his encouragement has allowed me to achieve so much more than I thought I could. I would like to thank Dr Francisco Ortega and Dr Indrakshi Ray for all of their support in the time collaborating with them and their amazing labs.

Thank you to both Dr Jerald Thomas and Dr Anne Cleary for volunteering their time to be a part of my Master's advisory committee. Thank you to all of the professors who have supported me throughout my undergraduate and graduate degrees here at CSU.

Thank you to all of my friends and peers across the ARefin Lab, NUI Lab, and others for all of the conversations that informed and improved my research. Lastly, thank you to my friends and family who have been an amazing support system as I have worked on achieving my Master's degree.

DEDICATION

I would like to dedicate this thesis to my immediate family, Ann, John and Drew Sturgeon - who have supported me and pushed me every step of the way. I could not have done it without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction[1]

1.1 Augmented Reality

Augmented Reality (AR) is the visual process of displaying virtual information and objects, overlaid on the real world. Augmented Reality shows itself in many forms, but one of the most common methods of displaying augmented information is via head-mounted display (HMD). AR projects virtual objects into 3D space and displays them within the surrounding world. HMDs currently utilize two methods of rendering virtual information, one of these rendering methods has made significant developments in both the research and commercial sectors recently. This rendering method is called optical see-through (OST) and consists of a semitransparent optical mechanism which allows for a near perfect view of the real world, similar to glasses, while superimposing virtual information into the view. While other rendering methods exist, OST allows for sharp and accurate virtual information combined with the clear and fully accurate view of the real-world through clear lenses. As one may guess, the ability to add virtual information into a person's view can allow for various new opportunities.

1.2 Applications of Augmented Reality

The main ability of AR is aid the user in the completion of a variety of real world tasks by giving them additional information in the form of virtual overlays and information. This general use of AR allows for applications across various fields, including education, medicine, military, and more [4–6]. Here are some specific examples of this use:

---

[1] Matthew Sturgeon, Dakota Kenoyer-Healy, Russel Cohen Hoffing, Steven Thurman, and Mohammed Safayet Arefin. Examining eye vergence during perceived depth changes with eye tracking system in optical see-through augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Daejeon, South Korea, 2025. IEEE.

Medical Sector: Medical sector jobs can utilize AR by utilizing virtual displays to aid and inform surgeons for more efficient surgeries. Azuma et al. has revealed that AR allows for real time rendering of 3D datasets, such as MRI scans, CT scans and ultrasounds, into the surgical space. This additional information utilizes methods known as "X-Ray Vision" where virtual information is displayed behind non-transparent real world objects. These methods could aid in advancing minimally invasive surgeries to allow for less cuts as well as smaller cuts for the patient [2].

Navigation: Augmented Reality has been tested and applied for to aid in navigation all across indoor and outdoor spaces within the real world. Additional information can be integrated into visual displays to direct the user to their objective. This has been done for indoor spaces with a hand camera, displaying an augmented video stream using fiducial markers for position tracking [7]. Similarly, augmented reality has been utilized to provide the user's precise location as well as the ideal route to several different destinations in outdoor spaces [8, 9].

Manufacturing and Maintenance: Traditionally, within manufacturing and maintenance work, workers are trained using text and images before eventually gaining access to hands on training and work. While this is effective, AR can give trainees opportunities to view and interact with 3D models of the equipment, giving a more realistic training experience [2]. In a similar vein, Doshi et al. utilized projector-based AR systems during a manual spot welding task, improving both the precision and accuracy of the weld. [10].

1.3 Problem Statement

While the wide utility of AR is exciting, the use of this new technology to its full potential is not quite possible yet. The process of displaying and utilizing information correctly is far more complex, as with the display of virtual information in space, comes the importance of depth and depth perception. For example, consider a surgeon who uses an OST AR device to visualize the depth of a tumor during its removal. The surgeon must accurately perceive the depth of the AR information and adjust their eye focus from the tumor to other surgery-related information at different depths to complete the surgery successfully. Thus, accurate and precise perception of AR depth is crucial.

Despite it's importance, depth perception has been a continued problem within AR depth perception research with several publications finding underestimation or overestimation of virtual objects compared to real objects [11–15]. These studies continue to research and theorize what causes are AR depth perception problems, hoping to better understand and improve depth perception within AR HMDs. Field of view, lack of depth and environmental cues, and Vergence Accommodation Conflict (VAC) have become some of the prominent causes of AR underestimation and overestimation. Most of the existing knowledge on depth perception of AR objects is based on subjective verbal reports and perceptual action-based tasks that depend on individual cognitive and perceptual abilities. Thus, depth dependent components of the human visual system during depth perception of AR objects, such as VAC, have not been thoroughly investigated. The incorporation of eye trackers into commercial OST devices, such as the Microsoft HoloLens 2 or Magic Leap 2, allows researchers to explore new research questions related to human visual system properties in OST AR.

Two key oculomotor processes that significantly influence perceptual focus on real and AR objects binocularly are *accommodation* and *vergence*. The processes of accommodation and vergence are interconnected, and this relationship is known as the *vergence-accommodation reflex* [16, 17]. Each has independent cues which drive the process. During eye *accommodation*, eye lens changes the shape to enable seeing information in sharp focus. The primary stimulus driving accommodation is blur. This is a monocular adjustment that each eye performs independently of each other. Besides, viewing objects at different depths requires the eyes to rotate, which is known as *vergence*. The primary stimulus that initiates vergence is binocular disparity, which involves rotating the eyes until the images of the left and right eye fuse. In binocular vision, the horizontal rotations of the eyes are used to focus on objects at varying distances. This action creates an angle between the eyes' visual axes, known as the *eye vergence angle (EVA)* [18–21] (Fig. 2.4). There are two types of EVA movements: *convergence* (when the eyes move from looking at a distant object to a nearby one) and *divergence* (when the eyes move from a close object to a farther one). Accordingly, nearer objects are associated with a higher EVA and the farther objects are associated with a smaller EVA [20].

Previously, Arefin et al. [19] investigated the relationship between EVA and depth of virtual stimuli in the virtual reality (VR) environment. The lack of a ground truth control in their study constrained the generalizability of the findings. A subsequent study by Arefin et al. [21, 22] involved a controlled experiment to evaluate EVA for targets in three environments: real, AR, and VR. The experiment employed a four-alternative forced-choice task, with participants wearing the HoloLens 2. Eye-tracking data were collected using a Pupil Labs eye tracker, custom-mounted on the HoloLens 2. The findings revealed the expected pattern of relationship between depth and relative changes in EVA, but showed significant individual variation in baseline EVA measurements. Authors suggested that this unexplained variability could be attributed to factors such as device properties, eye tracker accuracy and calibration, IPD, eye camera positioning, and other. The authors noted that their EVA results might be particular to the Pupil Labs eye tracker, rather than applicable to various OST AR devices and eye trackers. Thus, the lack of generalizability of EVA among individuals, OST AR devices, and eye-tracking systems highlights a significant literature gap.

The aim of the present study is to examine EVA associated with perceptual depth changes in two commercially available OST AR devices, to *confirm* and *generalize* the EVA mechanism across individuals, OST AR devices, and eye trackers. Our primary contribution encompasses two main aspects of experimental research: **replication** and **innovation**. We conducted a partial replication (a partial replication involves repeating an original experiment with intentional minor modifications [23]) of the experiment reported by Arefin et al. [21,22] by adopting the experimental task, depths, and EVA calculation method. Our replication confirms Arefin et al.'s [21] findings on the EVA pattern linking depth to relative changes in OST AR systems. Arefin et al. [21] noted unexplained individual variability in EVA and hypothesized it was due to the external eye tracker module attached to the HoloLens 2. We employed the built-in eye trackers of the OST AR devices, allowing us to assess whether individual variability was due to the embedded eye tracker. Overall, our replication research contributes to the long-standing "*Replication Crisis in AR-VR*" concern of the AR-VR research community through a successful partial replication experiment.

In addition to the partial replication, we used two commercially available single-plane OST AR devices (Microsoft HoloLens 2 and Magic Leap 2). The devices differ in focal plane positions, FOV, resolution, refresh rate, weight, and additional aspects (Section 3.2 for details). Despite substantial advancements in AR, the availability of OST AR devices equipped with eye-tracking technology remains limited. Microsoft HoloLens 2 (HL2) and Magic Leap 2 (ML2) are among the high-end OST AR devices available, featuring integrated eye-tracking functionalities necessary for this experiment. Therefore, we leveraged the built-in eye-tracking systems of the OST AR devices instead of external modules. Eye trackers in both HL2 and ML2 differ in the number of eye cameras, calibration methods, and accuracy (detailed in Section 3.2) allowing us to gain insight into variability of EVA. Moreover, we deployed a systematic approach to gather verbal reports by presenting a single stimulus at a random depth per trial, in contrast to Arefin et al. [21, 22], where all stimuli were displayed at every depth per trial, potentially influencing perceptual bias in verbal estimates.

1.4 Research Goals

To our knowledge, this experiment is the **first** in the OST AR research domain to examine EVA responses to AR targets and matched real-world targets utilizing OST AR devices with integrated eye-tracker, while also using verbal reports for direct comparison. Therefore, this paper presents new insights by comparing EVA with depth changes across individuals, OST AR devices, and eye-tracking systems, beyond replication. We formulated the following research questions.

RQ1: Are objective (EVA) and subjective (Verbal Estimate) measurements of depth perception consistent with depth of real and AR objects?

RQ2: Are objective measurements of vergence generalizable across OST AR devices and individuals?

Exploring these questions can help us better understand how EVA varies with perceptual depth changes, as well as the differences between subjective verbal estimates and objective EVA in depth perception under OST AR systems. Our research findings have the potential to contribute to the development of next-generation gaze depth-enabled adaptive OST AR systems. Previous research found that continuously switching eye focus from one depth to another depth (Focus depth switching) to integrate information decreased human performance and increased eye fatigue [24, 25]. For instance, when driving with an AR display, users must shift their focus from the vehicle ahead to AR information, resulting in *focus depth switching*. If EVA changes can represent focus depth switching, the AR display could render the information at the correct depth, reducing the need for focus shifts.

1.5 Conference Paper

I would like to acknowledge that a significant portion of this thesis is reused from the conference paper I published with my coauthors to IEEE ISMAR 2025 [26].

Chapter 2

Background and Related works[2]

2.1 Human Vision

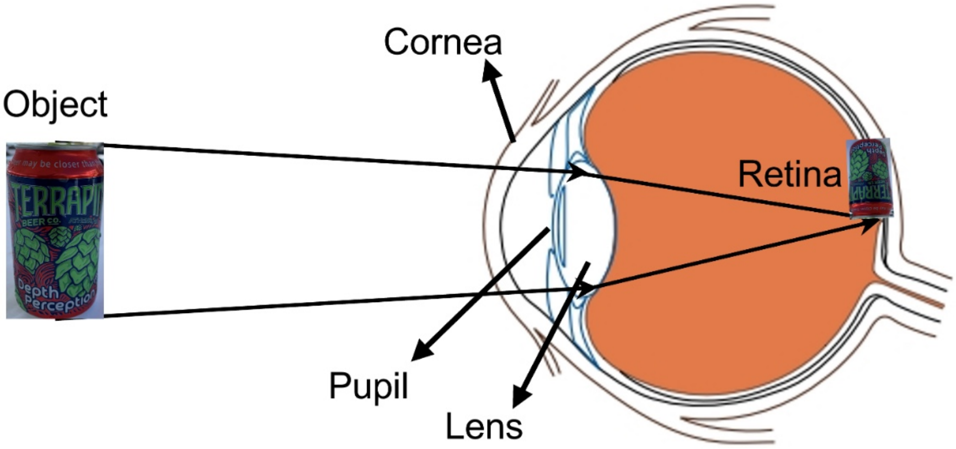


Figure 2.1: Diagram of vision in the retina [1].

The human visual system allows us as humans to take in external information and gain a 3-dimensional understanding of the external world. Several prior works have attempted to understand human perception and sensation, such as Fechner et al. [27]. As one might imagine, the visual process is extremely complex with several working muscles and fibers centered around the functionality of the eye. The eye functions as the input, gathering light on the retina to form an image of the external world as shown in Fig 2.1. The pupil is the black opening in the eye which allows controlled amounts of light to pass through. The pupil opens and closes to allow more or

---

[2] Matthew Sturgeon, Dakota Kenoyer-Healy, Russel Cohen Hoffing, Steven Thurman, and Mohammed Safayet Arefin. Examining eye vergence during perceived depth changes with eye tracking system in optical see-through augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Daejeon, South Korea, 2025. IEEE.

less light in as to not overexpose the retina to light. The diameter of the pupil in adults can range from 4 to 8mm in low light conditions, and 2 to 4mm in high light conditions [20, 28].

The visual system contains a lens located right behind the pupil. This lens is attached to the inside of the eye by zonule fibers and the ciliary muscle. The ciliary muscle pulls and pushes the lens into the correct shape to perfectly bend the light to hit the retina directly as in Fig 2.1. This manipulation of the light allows for a perfectly exposed and clear image of the external world to hit the retina upside down. The image is then flipped right side up by our brains and then correctly viewed [20].

2.2 AR Displays

As introduced previously, Augmented Reality displays have several different techniques of layering virtual information over the view of the real world. The two most common approaches in commercial devices being **Optical See-Through (OST)** and **Video See-Through (VST)**.

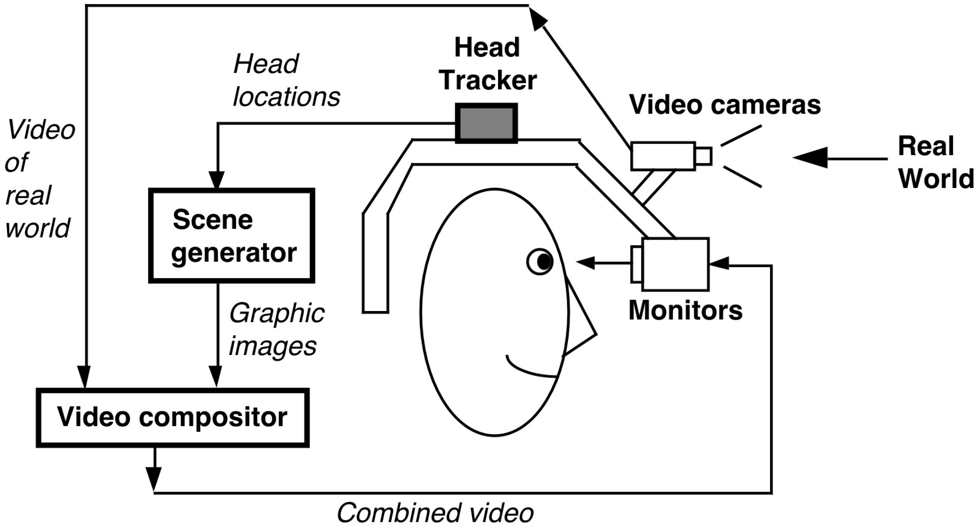


Figure 2.2: Diagram of a standard VST display [2].

Video See-Through Video See-Through displays use a similar technique as Virtual Reality by using monitors inside of the HMD to display both virtual and real information. The real world is displayed on the monitors using the video feed of external cameras mounted on the outside of the display with virtual information overlay-ed on top [29] as seen in Fig 2.2. These HMDs utilize sensors to track the headset's location and orientation with precision to ensure that the virtual information is rendered correctly [2].

Optical See-Through Optical See-Through displays on the other hand use an entirely unique technique for displaying virtual information. These headsets work similar to a thick pair of glasses to where the real world is directly visible through layers of glass or plastic called optical combiners. The virtual information is displayed in monitors above the optical combiners and then reflected into the eye by on top of the real information [29], resulting in Augmented Reality as seen in Fig 2.3. Similarly to VST displays, OST displays use a system of sensors to ensure that virtual information is displayed in the correct location [2].

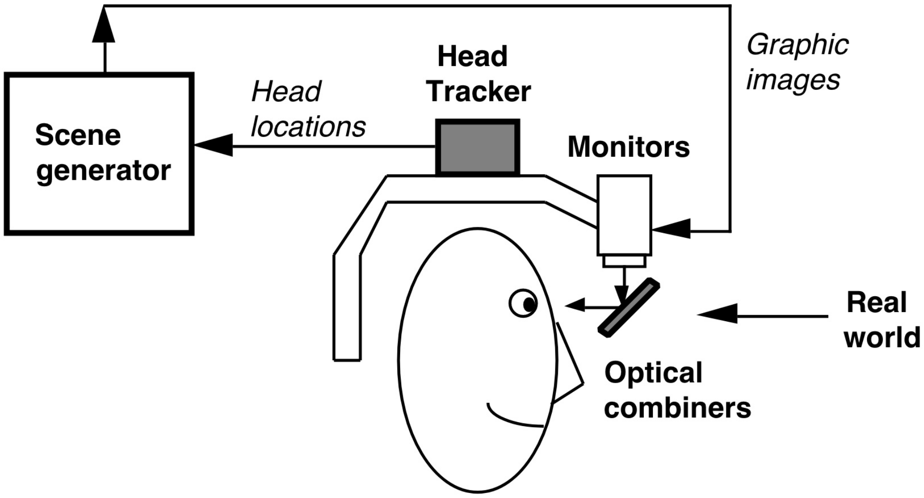


Figure 2.3: Diagram of a standard OST display [2].

2.3 Accommodation and Vergence

In natural viewing of real objects, the demands for vergence and accommodation are the same, and the vergence-accommodation reflex changes according to our visual system (left panel of

Fig. 2.4). However, viewing an AR object through the single focal plane OST AR system is different (right panel of Fig. 2.4). In a single-focal-plane OST AR system (e.g., HoloLens 2, Magic Leap 2, and others), the accommodation demand is the focal distance of the device's optical lens system. AR objects can appear in front and behind the display focal plane (see the right panel of Fig. 2.4)). Therefore, to see the AR object, the visual system needs to override the *vergence-accommodation reflex* to avoid double vision, and the vergence demand is not equal to the accommodation demand. This situation causes the *vergence-accommodation conflict (VAC)* [30, 31]. This conflict could potentially cause inaccurate depth and size perception, double vision, visual discomfort, and other consequences [16, 32, 33]. Although many such consequences of VAC exist, Vienne et al. [34] found that in stereoscopic displays, the vergence accuracy on the target itself is not significantly different between VAC and non-VAC conditions. Moreover, they also demonstrated that a larger VAC is not significantly different from a smaller one in terms of vergence accuracy.

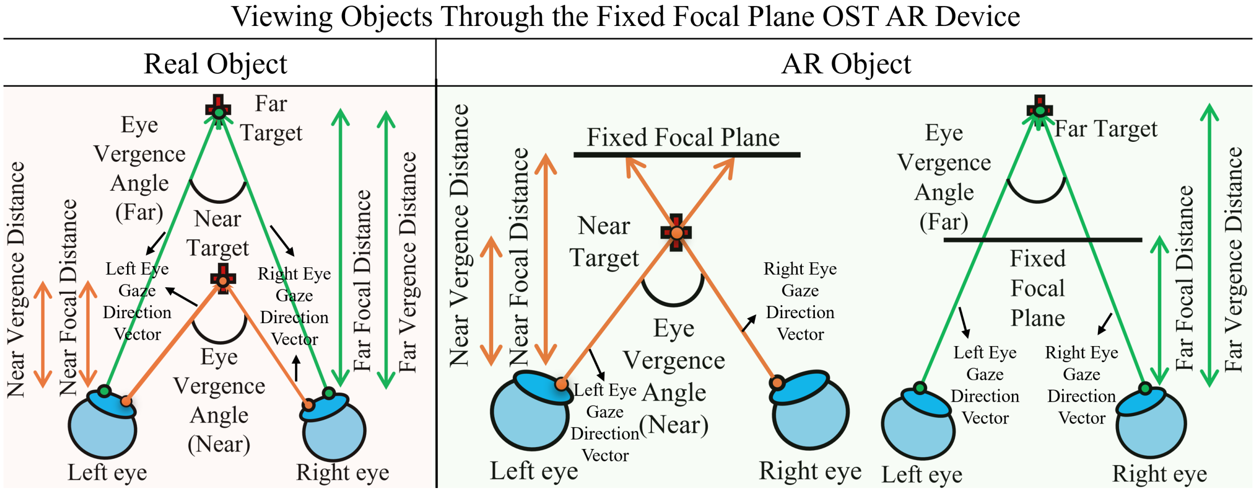


Figure 2.4: The vergence-accommodation conflict (VAC). In natural viewing (real object), vergence demand and accommodation demand are the same; hence, no VAC. VAC occurs in AR when accommodation is fixed at the display's focal plane.

Accommodation is typically measured using an infrared optometer to measure changes in the lens [35, 36]. However, some researchers have recently tried to use Purkinje reflections, which are reflections of light from the lens [37, 38]. Measurement of eye vergence angle has traditionally been measured using an eye tracker in a controlled experimental setup to understand the human visual system, VAC, visual attention, size perception, the reliability of the eye tracking systems, and other related concepts [34, 39–42]. Furthermore, AR and VR researchers have measured EVA while investigating changes in perceptual depth [19, 22], 3D gaze analysis [43], comparing gaze depth between real and virtual objects [44], effect of VAC [18], and others. There are traditionally two ways to measure vergence using an eye tracker, *the gaze-ray intercept model* and the *3D eye gaze vector model*. The gaze ray intercept model is created by calculating two gaze rays from the direction of the eyes. These rays rarely intersect, meaning that it is necessary to find the shortest line connecting the two rays and use the midpoint of this shortest line as a reference for gaze estimation. This additional step was added to improve the previous method, which used a plane to intersect with the rays [43]. The eye gaze vector model creates a 3D vector for each eye and uses an equation to create a "near-intersection point" to calculate vergence [45].

2.4 Depth Perception

Depth perception measuring techniques can be divided into three categories: verbal report, closed loop action based tasks, and open-loop action based tasks. Verbal report is measured by having the participant verbally report the distance they believe is between the object of measurement and themselves [11, 12]. Verbal estimates tend to be inaccurate as they require the participant to understand the measurement system used to get an accurate answer. In closed loop action-based tasks, the observer receives visual feedback while perceiving distance to an object, such as perceptual matching [15, 46], visual directed pointing [47], and others. In open-loop action-based task, the observer does not receive visual feedback while perceiving distance to an object, such as blind walking, blind matching, and others [12, 48–50]. Previous studies have shown that observers often make depth judgments that either underestimate or overestimate the depth of AR objects than real objects [11, 14, 51, 52]. Misperceptions regarding egocentric distances may result from factors including VAC, FOV, and the lack of depth and environmental cues, among others.

While cognitive and perceptual depth judgment techniques are common in the literature, there is significantly less focus on objective evaluations of ocular behaviors, including vergence, interpupillary distance, and pupil size, alongside perceptual and cognitive judgment methods. Ellis and Bucher [53] employed the perceptual matching task using AR objects superimposed on a physical background or occluder. Based on our knowledge, this is the only study that evaluated perceptual depth error with convergence and divergence. Moreover, rather than using an eye-tracker to measure EVA, they used a ratio between convergent and divergent eye movements during the appearance of an AR object. A prior study by Arefin et al. [19] investigated the relationship between EVA and the depth of virtual stimuli within a VR setting, demonstrating that EVA effectively tracks perceptual depth changes using the built-in eye tracker of the commercial VR display. Arefin et al. [21, 22] also explored EVA in real, AR, and VR utilizing HoloLens 2 and the Pupil Labs eye tracker. The research also involved measuring the verbal estimation of stimuli. The findings indicated variability in individual EVA, with EVA proving to be more accurate than verbal estimates.

Therefore, my research is the first to consider the integrated eye-tracking technology of OST AR devices to measure the depth-dependent component (vergence) of the human vision responses to AR targets and carefully matched real-world targets, while also leveraging verbal reports of the same stimuli for direct comparison.

Chapter 3

Experiment[3]

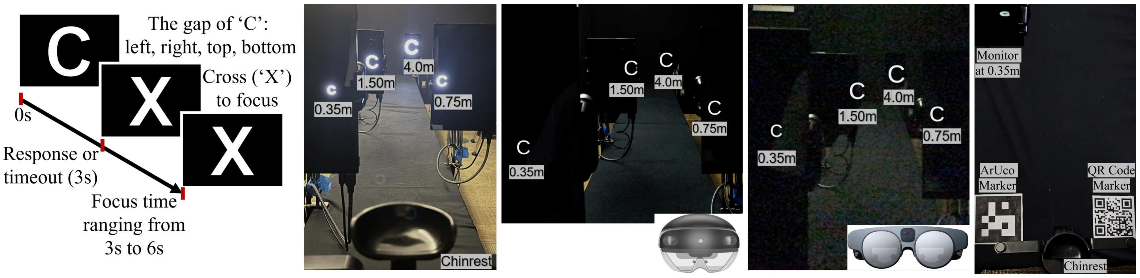
The aim of this experiment is to analyze the individual eye movements while utilizing real time eye tracking to calculate eye vergence angle in OST AR. This is a partial replication of prior research done by Arefin et al. [21,22]. Results will be compared against the original work to test for consistency. In this experiment we aim to assess whether objective (EVA) and subjective (Verbal Estimate) measurements of depth perception are consistent with the actual depth of real and AR objects. We also will be testing whether objective measurements of vergence are generalizable across OST AR devices as well as individuals. To properly assess this information both objective and subjective measurements will be gathered during the completion of the task via each device. The following chapter details the methods, variables and experimental setup which was used to accomplish these goals.

3.1 Experimental Task

The experimental task was adapted from the previous work of Arefin et al. [21, 22]. The rationale for selecting this task is that it ensures that the participants continuously focus on depth changes. The primary goal of the experimental task was to collect quantitative eye data for the EVA calculation at varying depths. To verify the participant's fixation at the intended depth, we employed a four-option forced-choice visual discrimination task. Each experimental trial began with a single uppercase letter C in a sans-serif font (Arial) displayed at a certain depth. Throughout the task, the character C was shown in a way that its gap could be oriented upward, downward, leftward, or rightward. The direction of C's gap was randomly selected for each trial. Participants were required to identify the direction and press the arrow key on the keyboard corresponding to the

---

[3] Matthew Sturgeon, Dakota Kenoyer-Healy, Russel Cohen Hoffing, Steven Thurman, and Mohammed Safayet Arefin. Examining eye vergence during perceived depth changes with eye tracking system in optical see-through augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Daejeon, South Korea, 2025. IEEE.



a) b) c) d) e)

Figure 3.1: (a) The four-alternative forced-choice visual discrimination task. (b) Real stimuli (e.g., 'C') were displayed on the physical monitors at four different depths. Participants viewed the real stimuli using the OST AR device; however, no AR information was displayed. (c) AR stimuli (e.g., 'C') were presented through Microsoft HoloLens 2 at the same depths and location as the real stimuli. (d) AR stimuli (e.g., 'C') were presented through the Magic Leap 2 at the same depths and locations as the real stimuli. Note that, for the sake of stimulus visualization in the image, stimuli were displayed at four different depths simultaneously. However, during the experiment, only one stimulus was displayed at a specific depth, as required by the task. A constant visual angle of 2.3°was used to ensure the real and AR stimuli were the same size at different depths. Note that camera-captured images do not represent the genuine perceptual experience of the experimental stimuli as observed by the human eye. (e) ArUco marker for Magic Leap 2 and for HoloLens 2, a QR Code was used to calculate the precise position of the AR stimuli.

orientation of the stimuli, all within a 3-second time frame. Following the participant's response or a period of 3 seconds, the 'C' was replaced by a cross ('X') at the same depth and location. The participants were then instructed to maintain focus on the cross ('X') for a random duration between 3 and 6 seconds. The sequence of changes in stimuli is shown in Fig. 3.1a. During the experiment, participants viewed a single piece of information at a particular depth at any given moment.

3.2 Apparatus and Setup

3.2.1 Physical Monitor:

Four identical physical monitors were used to display stimuli at four different depths. Each physical monitor, produced by the company *Bimawen*, featured a 4K OLED display with a resolution of 3840×2160, a 13.3-inch diagonal screen size, a 16:9 aspect ratio, 60Hz refresh rate, 400 cd/m^2 brightness, adjustable angles from 0°to 90°, and a weight of 750g. Each monitor was mounted on a fully adjustable 360°monitor holding arm, attached to the edge of the long table using a clamp base (see Fig. 3.1b). The four monitors did not overlap.

3.2.2 OST AR Device:

To present the AR stimuli in our experiment, we considered two commercially available OST AR devices: Microsoft HoloLens 2 and Magic Leap 2. The HoloLens 2 is an OST AR device with a single focal plane, featuring a focal distance of 2.0m, a diagonal FOV of 52°, a resolution of 1440×936, a refresh rate of 60Hz, and a weight of 566 grams. It monitors its position using 4 visible-light cameras [21, 54, 55]. The HoloLens 2 was used on medium brightness at 5/10. Magic Leap 2 is also an OST AR device with a single focal plane (focal distance of 0.74m), diagonal field of view (FOV) of 70°, resolution of 1440×1760, refresh rate of 120Hz, 16.8M color supported capability and weighs 260g [56, 57]. The Magic Leap 2 has a 12.6MP auto-focus RGB camera to scan information such as QR and barcode. The Magic Leap 2 was used in transparent display mode, which utilizes the *dynamic dimming feature* that adjusts the display's brightness based on real-world lighting. Brightness was difficult to manage with the Magic Leap 2 having the dynamic dimming feature which tries to optimize the view between real and AR objects which we wanted but because of this, we tried to make the brightness of the HoloLens 2 and the real condition as similar to the Magic Leap 2's brightness in the experimental environment as possible, but was not verified.

3.2.3 Eye Tracker:

HoloLens 2 and Magic Leap 2 are equipped with eye-tracking cameras that monitor both left and right eyes. The HoloLens 2 includes two infrared cameras, one per eye, which record eye movement data at 60 frames per second (fps). Conversely, Magic Leap 2 is designed with four eye cameras, two for each eye, that track eye movement data at a rate of 90 fps. We considered Microsoft's Mixed Reality Toolkit for Eye Tracking (MRTK2), specifically the extended eye tracking API of MRTK, to access eye movement data from the HoloLens 2 eye tracking system. To gather the eye movement data with Magic Leap 2, we used Magic Leap SDK. The HoloLens 2 eye tracking has an accuracy of about 1.5 degrees [58]. For Magic Leap 2, eye tracking has an accuracy of approximately 1 degree [59].

Table 3.1: Specification comparison between the two commercial OST AR devices and their built in eye trackers.

Device Specifications	Hololens 2	Magic Leap 2
Eye Tracking Accuracy	1.5°	1°
Eye Tracking Frame Rate	60 fps	90 fps
Eye Tracking Cameras	2 Eye Cameras	4 Eye Cameras
Diagonal Field of View	52°	70°
Resolution	1440 × 936	1440 × 1760
Refresh Rate	60Hz	120Hz
Focal Distance	2.0m	0.74m
Weight	566g	260g

3.2.4 Device Calibration:

To ensure accurate and precise eye tracking and AR rendering, we take into account the default calibration methods specific to HoloLens 2 and Magic Leap 2 devices. Although exact calibration methods are not disclosed, significant differences are apparent. Each device employs a customized calibration process that involves an eye-tracking calibration with small shapes placed at various positions. The primary differences are that Magic Leap 2 includes a specific fitting calibration and rejects calibrations if they are deemed invalid. Conversely, the HoloLens 2 lacks a fitting procedure and is generally less stringent about the eye calibration process, rarely, if ever, rejecting a calibration.

3.2.5 Real and AR Target Rendering:

Both the HoloLens 2 and Magic Leap 2 support marker tracking as a feature. In order to appropriately place the targets at the correct depths, we used the marker tracking features of these devices along with their spatial recognition features. For Magic Leap 2, an ArUco marker was used, and for HoloLens 2, a QR Code was used to calculate the precise position of the AR stimuli, matching those of the real-world stimuli based on the markers' positional data, which were fixed in position on the table (see Figure 3.1e). Earlier studies by Khan et al. [60, 61] applied the same rendering technique to display AR information in the real world environment, achieving alignment errors between the real and AR objects as minor as a few millimeters.

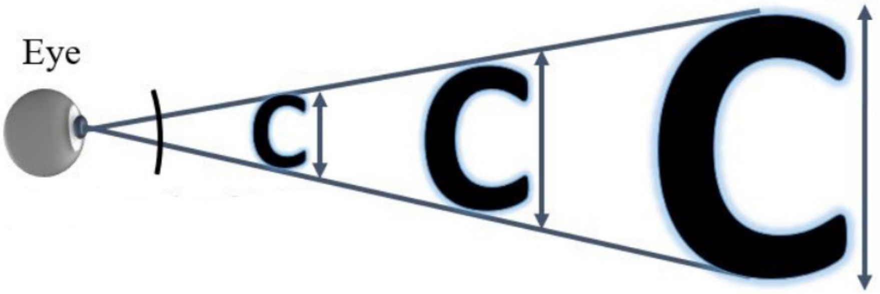


Figure 3.2: Diagram of visual angle adapted from [1]. The visual angle of 2.3°was used across the depths of 0.35, 0.75, 1.5 and 4 meters to have a corresponding virtual object size of 0.0141, 0.0301, 0.0602 and 0.1606 meters respectively.

Real and AR stimuli were displayed in the default bold format and rendered in white. A constant visual angle of 2.3°was used across all conditions (see in Fig 3.2) to ensure the real and AR stimuli were the same size at different depths, as established in prior research's experimental setups [24, 62, 63]. Fig. 3.1b, c, and d illustrate the real stimuli shown on the physical monitors as well as the AR stimuli viewed through the HoloLens 2 and Magic Leap 2, respectively.

3.2.6 Experimental Space:

The experiment took place in an entirely painted black room with no windows, ensuring the absence of external light influences as seen in Fig 3.3. A controlled and uniform lighting level of around $7 - 9 lux$ was maintained throughout the experiment.

3.2.7 Software:

To control real stimuli, a standalone Unity application (Unity version: 2021.3.4f1) was developed. Two separate unity applications were considered to control the AR stimuli, one for HoloLens

Figure 3.3: The experimental space contained a long table with a black cloth draped over to hold the monitors without distracting from the visual stimuli. Similarly, light was minimized in the space to allow navigation of the room while not distracting from the visual stimuli.

2 (Unity version: 2022.3.12f1) and another for the Magic Leap 2 (Unity version: 2022.3.12f1). Eye movement data pre-processing was executed with Python, while Rstudio was used for both visualization and statistical analysis.

3.3 Independent Variable:

There are three independent variables in this experiment: *environment*, *device*, and *depth*.

Environment (real and AR): In the real environment, target stimuli appeared on the physical monitor, whereas in the AR environment, they were displayed via the OST AR device. Participants wore the OST AR device for viewing both real and AR content binocularly. No information was presented through the OST device during the real environment condition.

Device (HoloLens 2, Magic Leap 2): When the device was HoloLens 2, the participants donned it, completed the real and AR environment conditions, and eye movement data were collected from the HoloLens 2 eye tracker. When the device was Magic Leap 2, the participants wore the Magic Leap 2, completed the conditions of the real and AR environment, and data on eye movement were collected using the Magic Leap 2 eye tracker.

Depth (0.35m, 0.75m, 1.5m, 4.0m): In our study, depth refers to the distance between the participant's eye position and the target stimuli. Four different depths were examined: 0.35m, 0.75m, 1.50m, and 4.0m. Given the properties of the near-clipping plane in the Magic Leap 2, AR content can be rendered as close as 0.35 meters. HoloLens 2 does not have these restrictions. Thus, for both the Magic Leap 2's rendering limit and typical close-range activities like reading with OST AR display, we used 0.35m. We considered 0.75m, as Jaschinski-Kruza et al. [64] indicated that the optimal comfortable viewing distance for human eyes is 51–99 cm, averaging 74 cm. Additionally, we drew inspiration from actions occurring within arm's reach, like engaging with AR interfaces. Inspired by action space distance activities, such as the optimal visibility distance for an automobile's head-up display (closer than 2.5m) [65, 66], we used a distance of 1.5m. Our study selected 4.0m, inspired by its application in navigation (e.g., HUD in aircraft) that is close to optical infinity [25]. According to Cutting and Vishton [67], accommodation and vergence are not effective depth cues at very long distances (e.g., $>$ 3m), even though they are measurable. Therefore, no further long distances were considered.

Participants shifted their focus from a starting depth to an ending depth in each trial, resulting in 12 potential start-end depth combinations (refer to Table 3.2), which we called *depth pairs*.

3.4 Dependent Variable:

We collected and documented eye movement data, task accuracy, and verbal response to perceived depth as our dependent variables.

Table 3.2: Twelve different combinations of start depth and end depth.

Depth Pair	Start Depth	End Depth	Depth Pair	Start Depth	End Depth
1	0.35m	0.75m	7	4.0 m	1.50m
2	0.35m	1.50m	8	4.0 m	0.75m
3	0.35m	4.0 m	9	4.0 m	0.35m
4	0.75m	1.50m	10	1.50m	0.75m
5	0.75m	4.0 m	11	1.50m	0.35m
6	1.50m	4.0 m	12	0.75m	0.35m

Eye Movement Data: From the HoloLens 2 eye tracker, we recorded timestamps, 3D eye position vectors, and 3D gaze direction vectors for the left and right eyes. Magic Leap 2 eye tracker recorded timestamps, data point confidence values, 3D eye center vectors, and 3D eye quaternion vectors for both the left and right eyes.

Accuracy: We collected user responses about the orientation of the gap in ’C’. If the user’s response aligned with the specified direction of C’s gap, it was considered correct; otherwise, it was deemed incorrect. The chance guessing rate for this task was 0.25 due to 4 alternatives (up, down, left, right).

Verbal Report: Participants verbally provided their subjective estimates of egocentric distances to real and AR stimuli.

3.5 Experimental Design

We employed a within-subject experimental design. We used a 2×2 Latin square to counterbalance the presentation order of the device and environment. Within each participant, each depth pair was shown in random order, with the constraint that the same depth could not be repeated consecutively. Each depth pair was repeated four times. Thus, each participant completed 2 (device)$\times 2$ (environment)$\times 12$ (depth pair)$\times 4$ (repetition) $= 192$ trials. At the beginning of each environmental condition (real or AR), the start depth of the first trial’s depth pair was randomly selected. In each subsequent trial, end depth of the current trial’s depth pair served as the start depth for the next trial's depth pair. For the verbal report, stimulus was presented at each depth randomly without repeating the same depth twice in a row. The stimulus was displayed at each depth three times. Each participant completed 2 (device) $\times$ 2 (environment) $\times$ 4 (depth) $\times$ 3 (repetition) $= 48$ trials for verbal reports.

3.6 Procedure

3.6.1 Pre-trial task:

At the beginning, participants were welcomed and given a brief overview of the experiment. Participants were required to carefully read the consent form and verbally confirm their agreement before proceeding. Following this, they completed a general questionnaire that encompassed demographic details along with an initial symptom questionnaire. Next, their interpupillary distance was measured at infinity with a commercial pupilometer, and eye dominance was assessed using the Miles test [68]. The experimenter then provided training on the experimental task using a physical training sheet, continuing until the experimenter was satisfied that the participant was proficient in the task. The physical training sheet contained examples of the stimuli. Throughout the training period, participants were instructed to ask any questions to clarify their understanding of the experimental task. Afterward, depending on the experimental condition, participants wore the OST AR device with the help of the experimenter. For clarity, the OST AR device was worn for the real condition as well, turned on for eye tracking purposes, without AR content displayed. The participants then completed the calibration process as outlined by the device manufacturer's built-in procedure. Neither experimental device (HoloLens 2 and Magic Leap 2) possesses distinct calibration methods specifically for eye trackers; instead, the devices utilized their general calibration for the eye tracker function.

3.6.2 Experimental trial:

After completing the pre-trial tasks, participants sat on a chair to complete the experiment. Participants were then instructed to place their chin on the chin-rest and place their right hand on the keyboard to perform the task. Only for the AR condition, participants were then directed to focus on the ArUco marker or QR code placed on the table, depending on the device they were equipped with (see Fig. 3.1e). The device subsequently displayed four ’C’ stimuli at varying depths in front of the participants. They were then asked to verify if they could observe all four stimuli within the display’s FOV at different depths. If any stimuli were not visible or clearly displayed at different depth, the device tracking system was reset, and the AR stimuli rendering confirmation process was repeated. After this step, the experimental trial begins.

Participants completed the symptom questionnaire after completing all trials of each environment (real or AR) using a specific device (HoloLens 2 or Magic Leap 2). Once all the trials of both conditions (AR and VR) for a device were concluded, they were prompted to respond to an environmental questionnaire. Subsequently, this procedure was repeated for the next device (HoloLens 2 or Magic Leap 2), which included calibration and task execution for both environments (real and AR). It is important to note that although symptom and environmental questionnaires were used, they fall outside the scope of this paper.

3.6.3 Verbal Report:

The verbal report values were collected only after completing all trials for each environment (real or AR) of each device (HoloLens 2 or Magic Leap 2). Upon presentation of a stimulus (a capital letter C with a gap to the right), participants were required to verbally estimate the perceived distance of the stimulus from their own location. They were free to use any measurement unit with which they felt most at ease for this depth estimation. Only one stimulus at a particular depth was displayed at a time during the verbal report procedure.

3.6.4 Post-trial task:

Upon completion of all experimental trials, a post-experiment informal interview was conducted to gather participant’s experience about the experiment. Then, participants were thanked and compensated accordingly. The entire experiment lasted approximately 1 hour and 15 minutes.

3.7 Participants

We used the G*Power software to do an *a priori* ANOVA within-subjects power analysis in order to determine the required number of participants. For one group with $(2 \times 2 \times 4 \times 4)$ = 64 repeated measures, a power of 0.80 with a small-medium effect size $d = .15$ using conservative values for correlation among measures (.5) and non-sphericity correction (.5), 19 subjects minimum were needed [69]. We recruited 24 participants from the local university community. Due to technical difficulties with the OST AR device's eye tracking, we were unable to collect data from 4 participants. Therefore, the results of the study are based on data from 20 participants (15 male, 5 female). The mean age of participants 24 years (range 20-29 years old). Because the integrated eye trackers in the Magic Leap 2 device were unable to perform a precise calibration for people with glasses we only included participants who did not wear eyeglasses. All participants had normal or corrected normal vision with contacts. Each participant was compensated with a 20 USD digital gift card for complete participation.

Chapter 4

Results and Discussion[4]

In this chapter, we lay out the study's results and discuss what they mean and what their importance is. We discuss EVA calculation, data pre-processing and analysis of both objective measures (EVA values) and subjective measures (verbal report values).

4.1 EVA Calculation

We used the EVA calculation methods demonstrated by Arefin et al. [21, 22], which utilized the left and right eyes' 3D eye gaze direction vectors projected onto a plane to create an angle (see Fig. 2.4). To calculate EVA using Magic Leap 2 eye tracking data, eye positions and eye rotation quaternions (the unit in Unity that represents rotations) were converted into 3D gaze direction vectors, also known as gaze vectors. This was accomplished by creating a Unity object that stores information, including position and forward vectors corresponding to each eye. By applying the eye position data to the Unity object's position and applying the eye rotation quaternions to the object's forward vector, we obtained the 3D gaze direction vectors for the left and right eyes for Magic Leap 2. The HoloLens 2 directly returned the 3D gaze direction vectors for each eye, so no changes were necessary. EVA was calculated using the equation presented below, which involves the 3D gaze direction vectors of the individual eyes ($\mathbf{L}(x, y, z)$ and $\mathbf{R}(x, y, z)$).

$$\text{EVA} = \arccos\left(\frac{\mathbf{L}(x, y, z) \cdot \mathbf{R}(x, y, z)}{|\mathbf{L}(x, y, z)||\mathbf{R}(x, y, z)|}\right)$$

---

[4] Matthew Sturgeon, Dakota Kenoyer-Healy, Russel Cohen Hoffing, Steven Thurman, and Mohammed Safayet Arefin. Examining eye vergence during perceived depth changes with eye tracking system in optical see-through augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Daejeon, South Korea, 2025. IEEE.

4.2 Data Pre-Processing

EVA values were cleaned in three steps. **First**, we processed blinks and/or noise. For Magic Leap 2, we changed any EVA value with a confidence value of less than 0.75 to "Not a Number" (NaN) so it would not be considered in subsequent analyses and event-related averaging. Similarly for the HoloLens 2, values of zero (i.e. blinks or signal dropout) were all converted to NaN. In the **second** step, behaviorally implausible data was replaced with NaNs. Maximum eye saccade velocities are about 700 degrees per second, with an average of around 300-400 degrees per second [70]. Since our analysis focuses on periods of fixations, we used an intermediate threshold of 500 degrees per second as well as any velocities outside a standard deviation of 2.5 from the mean velocity. As a result of data cleaning, 4.46% of AR and 11.08% of the real EVA values were replaced with NaN. It is important to note that all NaN values were not included in the analysis and did not impact the results. The **third** step was to discard trials that had less than 50% of their data intact after the previous cleaning steps. After this cleaning, 99.39% of AR trials and 99.23% of real trials were considered valid, meaning 12 AR trials and 15 real trials were removed out of 1920 total trials for each environment. Because over 99% of the data was retained no further data cleaning was needed.

After cleaning, we only considered the EVA values after the participant gave his/her behavioral response regarding the direction in which letter C was facing. We used the time range of *500 to 2000 ms* after pressing the response button to calculate the overall mean for each trial. This time range allows us to map EVA to the depth corresponding to the stimulus on each trial. We also excluded the first depth trial as there isn't a known starting depth.

Additionally, we cleaned the 960 verbal report values provided by participants, due to the presence of extreme outliers skewing the data distribution. It is typical for verbal reports to exhibit some highly inaccurate values (outliers). We used an outlier normalization process to minimize the impact of outliers on the data. Outliers were identified using the Inter Quartile Ratio (IQR). To calculate the IQR, the data were divided into four equally sized groups, and the IQR represented the range of the central (second and third) groups within these data. The outliers were then identified by creating a range that spanned from the lower bound of $mean - (IQR * 1.5)$ to the upper bound of $mean + (IQR * 1.5)$. Any values outside this range were considered outliers, and were then adjusted to equal the median of the verbal report dataset.

4.3 Analysis

We performed repeated measures ANOVA analysis to statistically assess the effects associated with our experimental variables using the ez package in R [71]. Other researchers used a similar statistical method in depth perception when all experimental variables are within subject [12, 50]. In cases where significant main and interaction effects were found, a Bonferroni test was performed using the DescTools package in R [72]. For effect size, we report the generalized eta squared (η^2), with interpretations as follows: small ($\eta^2 = 0.01$), medium ($\eta^2 = 0.06$), and large ($\eta^2 = 0.14$) [73]. We considered $p < 0.05$ a standard measure of significance.

4.4 Eye Vergence Angle (EVA) and Verbal Report

4.4.1 Results

To compare depth and environment across devices, we first calculated the mean EVA and verbal depth report across all repetitions and expressed EVA in degrees. Conversely, participants gave their perceived depth verbally in their preferred units, which we subsequently converted into meters. The summary statistics of verbal report across all participants, environments, devices and repetitions are: *0.35m*: $M = 0.33, SD = 0.17$, *0.75m*: $M = 0.71, SD = 0.24$, *1.5m*: $M = 1.41, SD = 0.3$, *4m*: $M = 3.51, SD = 0.98$. It is necessary to transform both measurement units (degree and meter) into a common dimensionless scale to facilitate a direct comparison between the verbal report and EVA. To achieve this, we computed a logarithmic ratio (AR/real) for both EVA and verbal report mean data for each participant, using the real condition value as ground truth. As recommended by Bland and Altman [74], applying a natural logarithm transformation can normalize the data by symmetrically distributing them around zero, making it more appropriate for standard statistical analyses. Consequently, we applied a natural logarithm transformation to the log ratio results (AR/real) from both the EVA and verbal report data. Here, a ratio of 0 signifies that participants perceived the depth of the AR object similarly to the real object. A ratio greater than 0 suggests that the AR stimulus's depth was underestimated relative to the real stimulus, while a ratio less than 0 indicates an overestimation of the AR object relative to the real stimulus (Fig. 4.1).

Fig. 4.1 shows the direct comparison of EVA and verbal estimate depth measurement methods for the HoloLens 2 and the Magic Leap 2. When *device*, *depth judgement methods (EVA and verbal estimate)* and *end depth* conditions were compared against average log ratio (AR/real) values, three way ANOVA analysis revealed a significant main effect of end depth on log ratio values ($F_{1,19} = 8.25, p < .01, \eta^2 = 0.031,$ *small*). Subsequent Bonferroni post-hoc analysis indicated that the log ratio at a near depth (0.35 m) differs significantly from that at other depths, except at 0.75 m, when measured using both EVA and verbal estimates with HoloLens 2 and Magic Leap 2. We found no statistically significant main effect of the device ($F_{1,19} = 0.44, p = .52, \eta^2 = 0.0048$) and depth judgement methods ($F_{1,19} = 0.0076, p = .93, \eta^2 = 1.0e - 04$). No other statistically significant interaction effects were observed.

4.4.2 Discussion

Our objective was to assess the consistency between objective and subjective measurements of depth perception when considering real and AR objects (RQ1). More precisely, we investigated whether the EVA and verbal report methods yield comparable results for depth assessment of both real and AR objects utilizing HoloLens 2 and Magic Leap 2. Our statistical analysis did not find significant differences between the EVA and verbal estimate depth measurement techniques when using the HoloLens 2 and Magic Leap 2 devices. This contradicts the findings reported by Arefin et al. [21, 22], which observed that EVA was more accurate with log-ratio data compared to verbal report, where verbal report was significantly underestimated. One possible explanation is the methodological variation in stimulus presentation for verbal reports between our study and that of Arefin et al. [21, 22]. We randomly presented each stimulus at various depths, in contrast

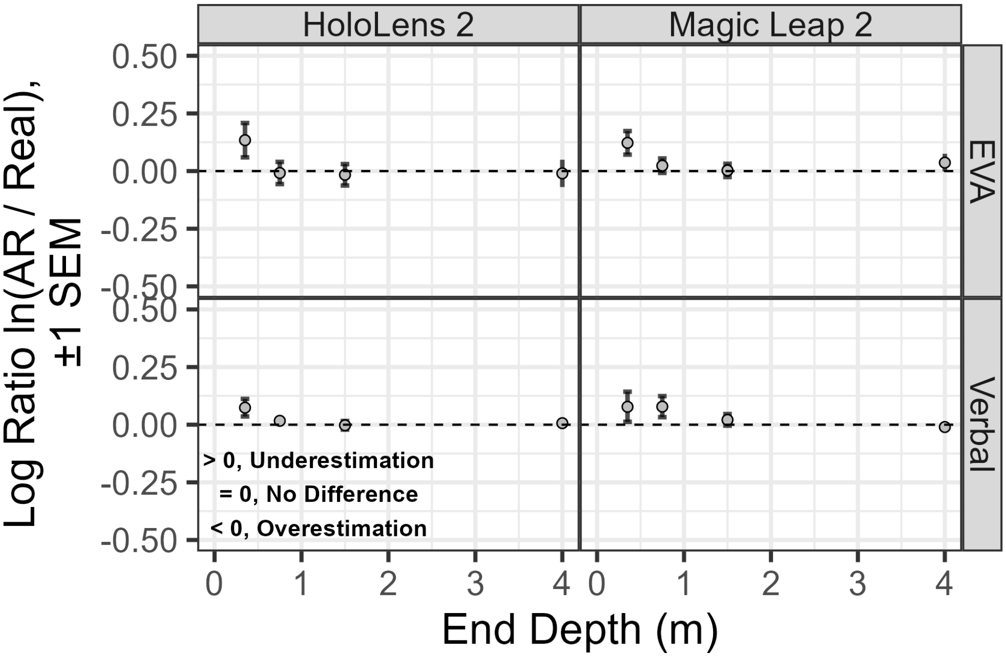


Figure 4.1: The verbal estimates and EVA using a natural log ratio of the AR object values divided by the real object values. These values are then plotted according to the corresponding device and the end depth at which the data was gathered in each panel. A value above zero on the Y axis indicates that the AR object was underestimated compared to the real object. A value below zero on the Y-axis shows that the AR object was overestimated compared to the real object.

to Arefin et al. [21, 22], where the authors simultaneously presented all stimuli at all depths while collecting verbal estimates. In general, our experimental results demonstrate that both objective (EVA) and subjective (Verbal Estimate) depth perception measurements align with the depths of both real and AR objects for HoloLens 2 and Magic Leap 2.

Our statistical analysis demonstrated a significant impact of end depth on the log ratio values for both the EVA and the verbal report. The subsequent post hoc test identified the end depth of 0.35 m as the source of significance. These results suggest that the AR object was underestimated relative to the real object exclusively at closer depth, as measured by both EVA and verbal report in HoloLens 2 and Magic Leap 2. Previous studies demonstrated an underestimate of the depth of AR objects compared to real objects in AR systems, as assessed through verbal reports [12, 50]. Contradicting these previous findings, our results, except at the closest depth of 0.35 m, indicate no such underestimation when using the HoloLens 2 and Magic Leap 2 devices. A potential reason could be that verbal reports are influenced by the cognitive skills of the participants. Consequently,

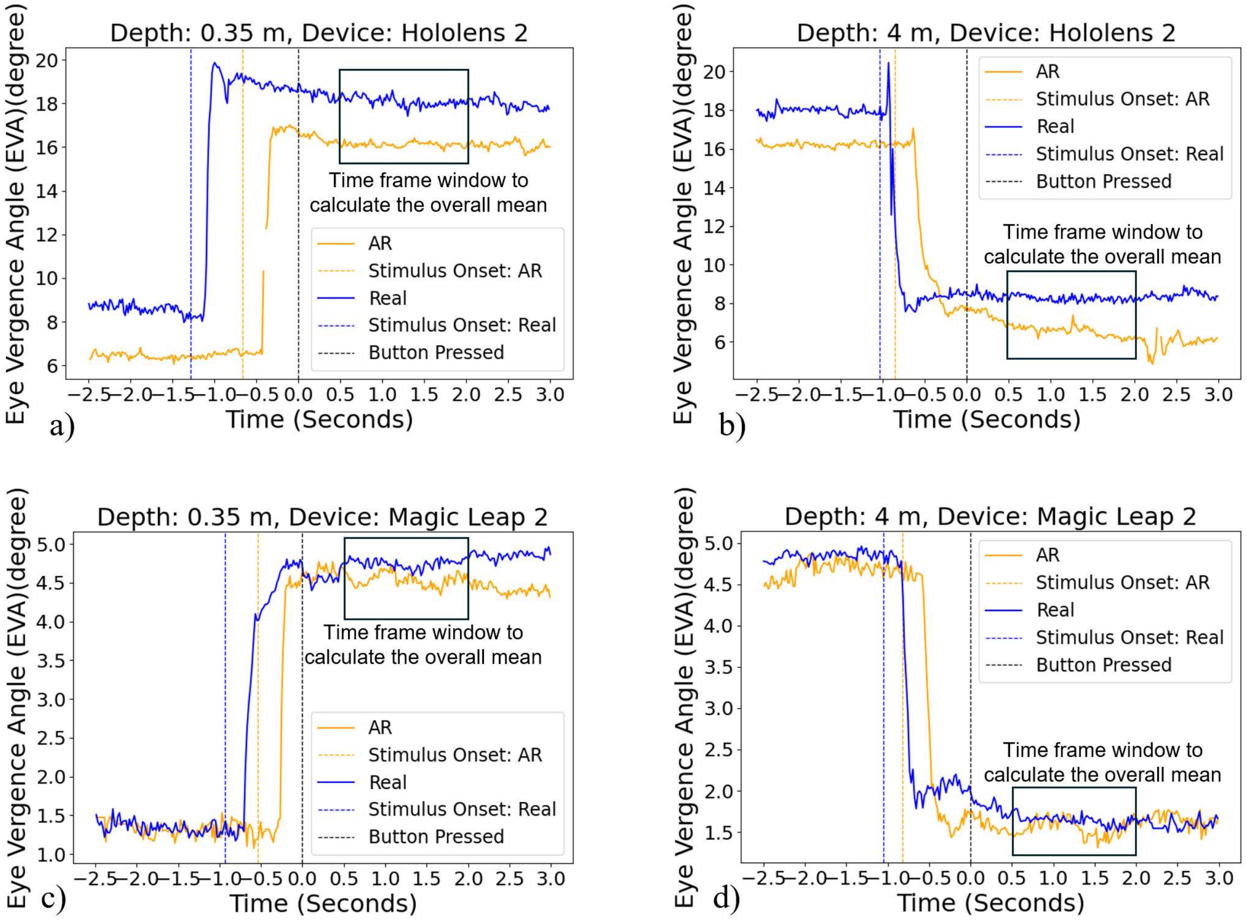


Figure 4.2: Two trials of unprocessed EVA data from same participant (P2) for both real and AR objects using HoloLens 2 and Magic Leap 2 are illustrated at two extreme depths in our experiment: 0.35m and 4.0m. In our study, stimulus onset refers to the moment the target, specifically the uppercase letter C, is presented during the task. Button-pressed indicates the participant's response to the direction of C's gap as per our task, confirming that the participant is focused on the target. We locked the button-pressed event to time zero. Therefore, the stimulus onset for both real and AR objects varied depending on when the button was pressed. EVA latency varies by approximately 200 milliseconds from the start of the stimulus for both real and AR objects, regardless of depth or device. The missing data (NAN) was coming from the AR condition with HoloLens 2 due to the unsuccessful tracking of eyes. With the use of integrated eye-tracking technology in OST AR devices, EVA effectively tracked the depth changes of real and AR objects following the stimulus onset.

replicating consistent results for depth perception in verbal reports remains a significant challenge because of cognitive variability of the participants. In alignment with earlier studies on the EVA of real and AR objects [21, 22], our findings indicate that the EVA values for AR objects closely match those of real objects at corresponding depths when using both the HoloLens 2 and Magic

Leap 2. A comprehensive discussion of the findings related to the real and AR objects' EVA is available in Section 4.5.2.

4.5 Eye Vergence Angle (EVA)

4.5.1 Results

Fig 4.2 presents an example that illustrates the changes in EVA values in two trials (real and AR) conducted with a participant (P2) at two extreme depths in our study: 0.35m and 4.0m, using HoloLens 2 and Magic Leap 2. Vision scientists have identified that vergence latency ranges from 160 to 260 milliseconds [75, 76]. Consistent with the existing literature, our findings indicate a comparable vergence latency (around 200 milliseconds) for both real and AR objects from the stimulus onset. This implies that vergence stabilizes at a new target depth following the vergence latency period. This further indicates that the current eye-tracking capabilities in the HoloLens 2 and Magic Leap 2 AR devices can effectively monitor relative changes in perceptual depth. This can potentially be achieved through the real-time computation of EVA as participants shift their focus when a target emerges. It should be noted that during data pre-processing, NaN values from HoloLens 2 were excluded from the analysis, thus not affecting the results.

For the EVA analysis, we calculated the mean EVA over participants, repetitions, environments, devices, and end depths. Fig. 4.3 shows the non-linear relationship between the EVA and end depth with real and AR objects for the HoloLens 2 and Magic Leap 2. This nonlinear relationship confirms that our measured EVA at different depths changes according to the binocular human visual system given the object at a fixed depth [20], the EVA decreases as the depth in meters increases. Furthermore, when *device*, *environment* and *end depth* conditions were compared, ANOVA analysis indicated a statistically significant main effect of end depth ($F_{1,19} = 95.71, p < .001, \eta^2 = 0.42$, *large*) and device ($F_{1,19} = 87.94, p < .001, \eta^2 = 0.74$, *large*). In addition, we observed a significant interaction effect between the device and the end depth ($F_{1,19} = 29.21, p < .001, \eta^2 = 0.082$, *medium*). This indicates that EVA differs significantly across end depths and devices. Bonferroni post-hoc tests revealed that statistically significant dif-

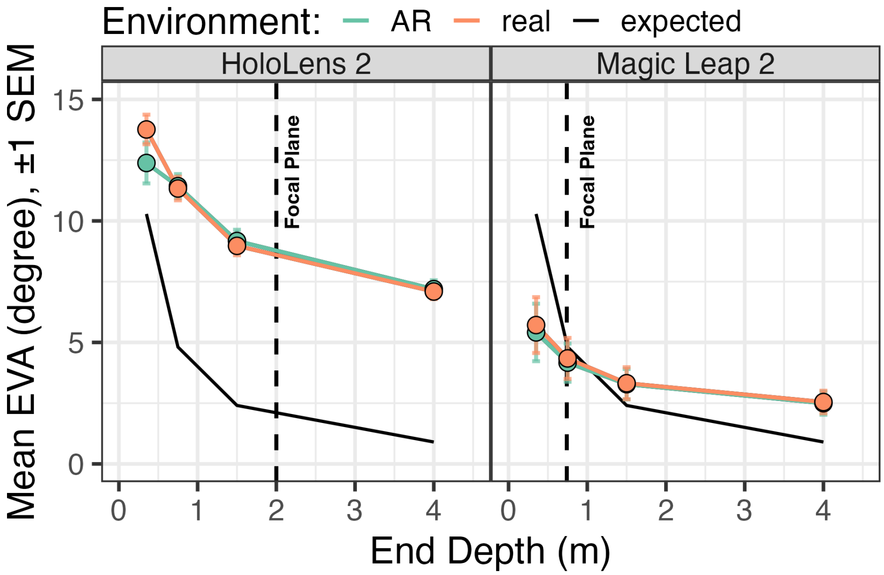


Figure 4.3: Results of EVA with end depth. Each point shows the mean of EVA for the time windows of 500ms to 2000ms after the button is pressed for each end depth. Each panel shows the results for each device and color-coded according to the environment. The dashed line shows the depth of the focal plane of HoloLens 2 and Magic Leap 2. EVA decreases with the increase of the end depth of the stimuli. EVA does not differ between the real and AR objects. For each end depth, EVA measurements varied significantly between the HoloLens 2 and Magic Leap 2. The solid black line represents the expected EVA based on the geometrical calculations of an individual with an average IPD of 63mm [3], if the stimulus is being displayed perpendicularly in front of the individual.

ferences were present across nearly all combinations of device and end depth, except for the pairs in Table 4.1.

Table 4.1: Pair of devices and depths that are not statistically significant

Pair	Device:Depth	Device:Depth
1	HoloLens 2: 0.35m	HoloLens 2: 0.75m
2	HoloLens 2: 1.50m	HoloLens 2: 4.0m
3	Magic Leap 2: 0.35m	Magic Leap 2: 0.75m
4	Magic Leap 2: 1.50m	Magic Leap 2: 4.0m
5	Magic Leap 2: 0.35m	HoloLens 2: 4.0m

It suggests that utilizing the HoloLens 2 and Magic Leap 2 eye trackers may present difficulties in employing EVA to accurately track depth changes within the range of 0.35m to 0.75m and 1.5m to 4.0m. Furthermore, no significant difference between the *Magic Leap 2: 0.35m vs. HoloLens 2: 4.0m* indicates a high variability of EVA between HoloLens 2 and Magic Leap 2. We found no statistically significant main effect of the environment ($F_{1,19} = 0.88, p = 0.36, \eta^2 = 0.0028$), indicating that EVA behaves similarly between the real and AR objects. No other statistically significant interaction effects were observed.

Fig. 4.4 shows individual panels for each participant with HoloLens 2 and Magic Leap 2, where the mean EVA was calculated over repetitions, environments, devices, and end depths for each participant separately. We determined the interquartile ratio (IQR) for each individual to quantify the level of per-participant variability. For this calculation, the EVA data of each participant was segmented into four groups of equal size, with the IQR reflecting the range between the central (second and third) quartiles of these data. For per-participant variability analysis, when *device*, *environment*, and *end depth* conditions were compared against IQR of EVA, ANOVA analysis showed a statistically significant main effect of device ($F_{1,19} = 16.56, p < .001, \eta^2 = 0.2019,$ *large*) and end depth ($F_{1,19} = 17.56, p < .001, \eta^2 = 0.1546,$ *large*). Moreover, we observed a significant interaction effect of device and end depth ($F_{1,19} = 5.80, p < .05, \eta^2 = 0.0366,$ *small*). To further investigate the significant interaction effect ($device \times enddepth$), we used a Bonferroni test to identify the factors driving the interaction effect. The findings showed a significant variation in the EVA IQR of HoloLens 2 participants at the end depth of 0.35m compared to all other depths, regardless of the device used.

4.5.2 Discussion

Our second research question (RQ2) seeks to answer whether EVA measurements are generalizable across devices and individuals.

Per-device Variability:

Our findings indicated a significant main effect of EVA by device and an interaction effect between device and end depth. There was no significant difference in EVA values with a target object at 0.35m on Magic Leap 2 and a target object at 4.0m on HoloLens 2. This suggests a significant variation in EVA between HoloLens 2 and Magic Leap 2. One potential explanation is the differences in eye tracking accuracy between HoloLens 2 and Magic Leap 2. The HoloLens 2 exhibits marginally lower accuracy (1.5°) compared to the Magic Leap 2, which has an accuracy of 1° [58, 59]. Additionally, Magic Leap 2 is equipped with four eye cameras (two dedicated to each eye), whereas HoloLens 2 utilizes two eye cameras (one assigned to each eye). Another possible reason could be that the devices employ different eye-tracking models for calibration. Our practical experience suggests that although both devices utilize eye-tracking-based calibration techniques, Magic Leap 2 uses a more rigorous calibration than HoloLens 2. For example, Magic Leap 2 disregards calibration efforts if it detects erroneous calibration points, such as when users wear eyeglasses. In contrast, HoloLens 2 consistently accepts calibration, regardless of the quality of the calibration. Unfortunately, the specific distinctions in terms of accuracy and algorithm between the eye-tracking calibration procedures of HoloLens 2 and Magic Leap 2 remain unknown due to the lack of open-source information. One potential cause of variability between devices is the difference in their fixed focal planes: HoloLens 2 (2.0m) and Magic Leap 2 (0.74m). As noted by Mon-Williams and Tresilian [77], human depth perception is biased towards the focal depth of the screen. Thus, targets in front of the fixed focal plane are overestimated (the estimated EVA is smaller than the expected vergence), while targets behind it are underestimated (the estimated EVA is larger). Fig. 4.3 illustrates the relationship between expected vergence (solid black line) and estimated EVA across devices. However, this theory aligns with the Magic Leap 2, unlike the HoloLens 2, as our method estimates device-specific EVA values within different ranges. Due to these potential factors, variability in EVA was observed between the devices. Our results show a non-linear relationship between depth and EVA (increasingly greater EVA for closer distances) that would be expected from previous research and our understanding of the geometry of the human visual system [20]. Thus, while *absolute* estimates of EVA do not generalize across devices, these results suggest that *relative* changes in EVA can be a useful measure for estimating perceptual depth with an appropriate baseline with respect to device and/or individual users.

Per-participant Variability:

Our results showed variability in EVA per participant between devices. Furthermore, we found significant differences in the association of per participant variability driven by EVA at the end depth of 0.35 m in HoloLens 2. However, we did not observe any per-participant variability of EVA in Magic Leap 2. We can see in the individual data that several participants (P3, P6, P9, P16, and P19) had a large (and unrealistic) drop in the EVA value for the closest depth (0.35m) in the HoloLens 2 AR data (see Fig. 4.4). We did not observe the same for Magic Leap 2 at the closest distance. The differences in EVA observed among participants suggest that objective measurements of vergence cannot be generalized across all individuals, supporting previous research findings [21,22]. Potential fundamental reasons for these findings include the position of the focal planes, algorithm limitations, FOV, and differences between the eye-tracking systems and calibrations of the OST AR devices. The near stimulus (0.35m) was far from the fixed focal plane (2m) of the HoloLens 2, but closer to the fixed focal plane (0.74m) of the Magic Leap 2. Therefore, participants observed the target at 0.35m with HoloLens 2 under large VAC compared to the Magic Leap 2. Therefore, this could potentially cause difficulties for the participant when focusing at the nearest depth with HoloLens 2. Our post-experiment informal interview with the participants further supports this. Some participants observed that it was occasionally challenging to maintain focus, resulting in double vision, when stimuli were presented at the nearest depth of 0.35m using the HoloLens 2. The experimenters sometimes experienced double vision when attempting to focus on this depth within the HoloLens 2, supporting this reasoning. Our EVA calculation utilized 3D gaze direction vectors from the eye tracker without checking gaze ray intersections. As mentioned before, double vision was observed at 0.35m with HoloLens 2 with some participants, indicating EVA might have been computed before eye fusion. Thus, algorithm limitations may influence per-participant variability. Another fundamental factor affecting per-participant variability is the field of view (FOV) of the devices; Magic Leap 2 has a 70° diagonal FOV, while HoloLens 2 has a 52° diagonal FOV. Consequently, participants using HoloLens 2 needed increased effort, such as more head rotation, to focus on the object at 0.35. A deficit in effort to focus on the target at the closest distance with HoloLens 2 might lead to EVA being inadvertently dropped. Additionally, as Section 4.5.2 outlines, varying cameras and calibration methods across devices could also affect this variability.

Real Object VS AR Object:

Lastly, our results showed no significant difference between real and AR objects across devices. It is noteworthy that in our experiment, we utilized two devices with fixed focal planes, allowing participants to view the AR object with the presence of vergence-accommodation conflict (VAC) (refer to Fig.2.4 and Fig.4.3). Although the purpose of our experiment was not to evaluate the impact of VAC on the EVA mechanism, we anticipated detecting its effects. Interestingly, we did not observe any impact of VAC on EVA in either device. An explanation for this could be found in the models of the EVA mechanism. The EVA mechanism comprises two parts: a fast component that alerts the eye to sudden depth changes and a feedback component that reduces vergence error for a specific depth through ongoing neurological feedback [34, 78]. Moreover, our findings on the impact of the environment are consistent with the study by Vienne et al. [34], where it was reported that the accuracy of the vergence angle when employing stereoscopic displays equipped with VAC does not vary substantially from that of a standard display or various VAC intensities. In Fig. 4.3, we notice that the curve of the EVA for real and AR conditions does match the expected trend of VAC for the Magic Leap 2, with the measured EVA appearing to be biased towards the focal plane instead of the anticipated values for a person with average IPD [77]. The HoloLens 2 does not directly demonstrate this behavior, but we do not have a matching depth for it's focal plane as a direct comparison. This may be explained by the differences in eye tracking and calibration between the two devices. A second possibility is that, due to the headsets being worn for the real condition to collect eye data, real condition could be potentially impacted by the focal plane of the device.

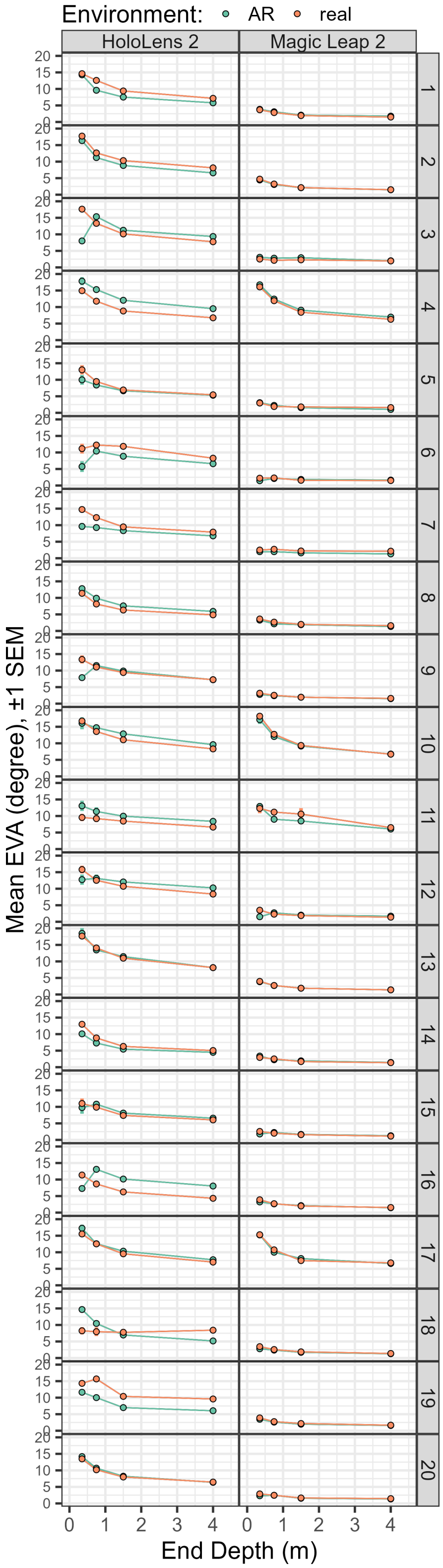


Figure 4.4: Individual graphs for each participant, illustrating mean EVA at each end depth for both environments on each device. The per-participant as well as per-display variability for EVA is visible.

Chapter 5

Limitations and Future Work[5]

Limitations of our research provide guidance for future work. Although these limitations do not undermine the significance and merit of our findings, they highlight the need for caution when interpreting our results in the context of practical applications and future replication efforts.

Participants Effect: Our findings are limited only to young adults of the university level. Additionally, we included individuals who did not wear eyeglasses due to the calibration challenges with Magic Leap 2. To generalize our findings to a wider range of people, future studies should consider middle-aged and older participants, as well as individuals with glasses, when examining EVA.

Perceptual Action Task: Our experiment did not consider any perceptual action-based depth perception task (e.g., perceptual matching, blind walking, etc.) along with the EVA measurements. Thus, a possible future investigation might integrate a perceptual action-based depth perception task, a verbal report, and EVA measurements to assess the depth of both real and AR objects.

Physical Monitor Effects: Due to the need to ensure that AR targets were correctly aligned, verbal AR reports may have been confounded with the physical location of the monitors. Therefore, a potential future work could replicate our experiment by separating the monitors from the AR stimulus.

VAC Effect: Our study was not designed to investigate the impact of different amounts of VAC effects on EVA. However, a larger difference between the focal plane and AR object's depth may

---

[5] Matthew Sturgeon, Dakota Kenoyer-Healy, Russel Cohen Hoffing, Steven Thurman, and Mohammed Safayet Arefin. Examining eye vergence during perceived depth changes with eye tracking system in optical see-through augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Daejeon, South Korea, 2025. IEEE.

impact the results. Therefore, a potential future study could investigate the positive and negative effects of VAC on EVA precision and latency in OST AR systems. The trend of the Magic Leap 2 in Fig. 4.3 suggests the necessity of using a consistent external eye tracker in studies to assess if the focal plane affects the EVA of real objects. This experiment should utilize one of the several devices which allows for direct control of the distance of the focal plane such as the AR Haploscope [79]. This would allow for control of positive and negative effects of VAC as well as control of VAC intensity.

Research Question: Does direct manipulation of the focal plane impact depth perception of a virtual object?

Same Device: An interesting future work could be to expand on the inconsistent tracking across device to see if objective measurements are consistent across two versions of the same device. Inconsistencies within manufacturing may even cause the resulting eye tracking and EVA values to not be generalizable.

Research Question: Are objective measurements of vergence generalizable across two iterations of the same AR Device?

Vergence Measurement Methods: In this study, one specific vergence measurement technique was used; however, several other measurement techniques exist which could have been used for this experiment. Future works could further inspect the pros and cons of each measurement technique and even compare their results with real data.

Research Question: Which vergence measurement technique is most effective in measuring the depth of the target object.

Motion Sickness: Motion sickness is another issue within mixed reality and virtual reality which causes difficulties in device usage. it is possible that VAC has an impact on motion sickness, a future study may use a similar device as to the haploscope or HMDs with different focal depths to study the impact of VAC on motion sickness using an SSQ questionnaire.

Research Question: Can VAC increase the effects of motion sickness?

Chapter 6

Conclusion[6]

Several different publications have reported over and underestimation in the depth estimation of virtual information in AR and VR with regards to depth estimation of real information. This causes several issues in AR as it can hinder the utility of the virtual information and the task which is being completed. This experiment examines an objective measure of depth perception through EVA and compares this to the subjective measure of verbal estimates and the display depth. Two different commercial devices were utilized along with their eye trackers for better generalizability across devices. Solving this research problem allows us a better understanding of how the human body interacts with AR and VR headsets. Eventually, this work could help lead to significant strides in correcting depth perception in AR and VR, allowing for better accuracy in displays.

To the best of our knowledge, this is the first systematically designed experiment to investigate how EVA behaves with depth changes using the integrated eye trackers of commercially available fixed focal plane-based OST AR devices. Additionally, we analyzed the subjective verbal estimation method for depth perception and compared it with objective EVA measurements. We experimented with two commercially available modern OST AR displays, HoloLens 2 and Magic Leap 2, using a four-alternative forced-choice visual discrimination task. We carefully calibrated the AR objects to appear in the same position as the real objects at four different depths: 0.35m, 0.75m, 1.50m, and 4.0m. One of the central goals of the eye-tracking-enabled OST AR system is to estimate the depth at which participants are looking and to provide information adaptively according to the perceptual depth. Our findings indicate that it is possible to estimate the user's perceptual depth by calculating the eye vergence angle (EVA) with the integrated eye tracker of the OST AR devices. Therefore, our findings will be essential for the advancement of next-generation

---

[6] Matthew Sturgeon, Dakota Kenoyer-Healy, Russel Cohen Hoffing, Steven Thurman, and Mohammed Safayet Arefin. Examining eye vergence during perceived depth changes with eye tracking system in optical see-through augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Daejeon, South Korea, 2025. IEEE.

adaptive gaze depth-based OST AR systems, which may significantly improve the user experience of OST AR systems. The primary findings of our research are listed below.

- The perceptual depth estimation with objective (EVA) and subjective (verbal report) measurements was consistent with the depth of real and AR objects.
- Our findings revealed a significant variation in EVA between HoloLens 2 and Magic Leap 2. Therefore, EVA can not be generalized across OST AR displays.
- Although the current eye-tracking capabilities in the HoloLens 2 and Magic Leap 2 AR devices can efficiently track changes in perceptual depth, we observed that EVA is ineffective for tracking depths beyond 1.5m.
- Our findings demonstrated variability in EVA per participant across devices. Consequently, objective vergence measurements cannot be generalized for all individuals.
- Our findings did not observe any significant differences in EVA between real and AR objects.

Bibliography

[1] Mohammed Safayet Arefin. Augmented reality fonts with enhanced out-of-focus text legibility. 2022.

[2] Ronald T Azuma. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4), August 1997.

[3] Neil A. Dodgson. Variation and extrema of human interpupillary distance. In Mark T. Bolas, Andrew J. Woods, John O. Merritt, and Stephen A. Benton, editors, *Stereoscopic Displays and Virtual Reality Systems XI*, volume 5291, pages 36 – 46. International Society for Optics and Photonics, SPIE, 2004.

[4] Mark Billinghurst, Adrian Clark, and Gun Lee. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2):73–272, 2015.

[5] Philip J. Edwards, Laura G. Johnson, David J. Hawkes, Michael R. Fenlon, Anthony J. Strong, and Michael J. Gleeson. Clinical experience and perception in stereo augmented reality surgical navigation. In Guang-Zhong Yang and Tian-Zi Jiang, editors, *Medical Imaging and Augmented Reality*, pages 369–376, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[6] W. Friedrich. ARVIKA-augmented reality for development, production and service. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pages 3–4. IEEE Comput. Soc, 2002.

[7] D.W.F. Van Krevelen and R. Poelman. A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Reality*, 9(2):1–20, January 2010.

[8] Wolfgang Narzt, Gustav Pomberger, Alois Ferscha, Dieter Kolb, Reiner Müller, Jan Wieghardt, Horst Hörtner, and Christopher Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, March 2006.

[9] Umair Rehman and Shi Cao. Augmented Reality-based Indoor Navigation: A Comparative Analysis of Handheld Devices vs. Google Glass.

[10] Ashish Doshi, Ross T. Smith, Bruce H. Thomas, and Con Bouras. Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. *The International Journal of Advanced Manufacturing Technology*, 89(5-8):1279–1293, March 2017.

[11] Bobby Bodenheimer, Haley Adams, Mirinda Whitaker, Jeanine Stefanucci, and Sarah Creem-Regehr. Perceiving absolute distance in augmented reality displays with realistic and non-realistic shadows. In *ACM Symposium on Applied Perception 2023*, pages 1–9. ACM, 2023.

[12] Holly C. Gagnon, Carlos Salas Rosales, Ryan Mileris, Jeanine K. Stefanucci, Sarah H. Creem-Regehr, and Robert E. Bodenheimer. Estimating distances in action space in augmented reality. *ACM Transactions on Applied Perception*, 18(2):1–16, 2021.

[13] V. Interrante, B. Ries, and L. Anderson. Distance perception in immersive virtual environments, revisited. In *IEEE Virtual Reality Conference (VR 2006)*, pages 3–10. IEEE, 2006.

[14] Joshua M. Knapp and Jack 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*, 13(5):572–577, 2004.

[15] Gurjot Singh, Stephen R. Ellis, and J. Edward Swan. The effect of focal distance, age, and brightness on near-field augmented reality depth matching. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1385–1398, 2020.

[16] Mark Mon-Williams and John P. Wann. Binocular virtual reality displays: When problems do and don't occur. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(1):42–49, 1998.

[17] V. V. Krishnan and Lawrence Stark. A heuristic model for the human vergence eye movement system. *IEEE Transactions on Biomedical Engineering*, BME-24(1):44–49, 1977.

[18] Julie Iskander, Mohammed Hossny, and Saeid Nahavandi. Using biomechanics to investigate the effect of VR on eye vergence system. *Applied Ergonomics*, 81:102883, 2019.

[19] Mohammed Safayet Arefin, J. Edward Swan II, Russell A. Cohen Hoffing, and Steven M. Thurman. Estimating perceptual depth changes with eye vergence and interpupillary distance using an eye tracker in virtual reality. In *2022 Symposium on Eye Tracking Research and Applications*, ETRA '22, New York, NY, USA, 2022. Association for Computing Machinery.

[20] Herbert Gross, Fritz Blechinger, and Bertram Achtner, editors. *Handbook of optical systems. 4: Survey of optical instruments*. Wiley-VCH, 1. repr edition, 2014.

[21] Mohammed Safayet Arefin, J. Edward Swan II, Russell Cohen Hoffing, and Steven Thurman. Mapping eye vergence angle to the depth of real and virtual objects as an objective measure of depth perception.

[22] Mohammed Safayet Arefin, J. Edward Swan II, Russell Cohen-Hoffing, and Steven M. Thurman. Tracking perceptual depth with eye vergence movements in real world, augmented reality, and virtual reality environments. In *Vision Sciences Society Annual Meeting Abstract (VSS), Journal of Vision*, volume 23, Aug 2023.

[23] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. Is once enough?: on the extent and content of replications in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3523–3532. ACM, 2014.

[24] Mohammed Safayet Arefin, Nate Phillips, Alexander Plopski, Joseph L. Gabbard, and J. Edward Swan II. The effect of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur on human performance in optical see-through

augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2014–2025, 2022.

[25] Joseph L. Gabbard, Divya Gupta Mehra, and J. Edward Swan. Effects of ar display context switching and focal distance switching on human performance. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2228–2241, 2019.

[26] Matthew Sturgeon, Dakota Kenoyer-Healy, Russel Cohen Hoffing, Steven Thurman, and Mohammed Safayet Arefin. Examining eye vergence during perceived depth changes with eye tracking system in optical see-through augmented reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Daejeon, South Korea, 2025. IEEE.

[27] Gustav Theodor Fechner. Elements of psychophysics, 1860. In Wayne Dennis, editor, *Readings in the History of Psychology*, pages 206–213. Appleton-Century-Crofts, New York, 1948.

[28] Pablo Artal. Image formation in the living human eye. *Annual Review of Vision Science*, 1(1):1–17, 2015.

[29] Kenneth R. Moser, Mohammed Safayet Arefin, and J. Edward Swan. Impact of alignment point distance and posture on spaam calibration of optical see-through head-mounted displays. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 21–30, 2018.

[30] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 2008.

[31] Gregory Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1912–1931, 2016.

[32] George-Alex Koulieris, Bee Bui, Martin S. Banks, and George Drettakis. Accommodation and comfort in head-mounted displays. *ACM Transactions on Graphics*, 36(4):1–11, 2017.

[33] Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3):30201–1–30201–14, 2009.

[34] Cyril Vienne, Laurent Sorin, Laurent Blondé, Quan Huynh-Thu, and Pascal Mamassian. Effect of the accommodation-vergence conflict on vergence eye movements. *Vision Research*, 100:124–133, 2014.

[35] Kenji Ibi. Characteristics of dynamic accommodation responses: comparison between the dominant and non-dominant eyes. *Ophthalmic and Physiological Optics*, 17(1):44–54, 1997.

[36] G Heron, W.N Charman, and C Schor. Dynamics of the accommodation response to abrupt changes in target vergence as a function of age. *Vision Research*, 41(4):507–519, 2001.

[37] Conny Lu, Praneeth Chakravarthula, Yujie Tao, Steven Chen, and Henry Fuchs. Improved vergence and accommodation via purkinje image tracking with multiple cameras for AR glasses. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 320–331. IEEE, 2020.

[38] Faik Ozan Ozhan, Ugur Aygun, Afsun Sahin, and Hakan Urey. Dynamic accommodation measurement using purkinje reflections and machine learning. *Scientific Reports*, 13(1):21625, 2023.

[39] Brian C. Daugherty, Andrew T. Duchowski, Donald H. House, and Celambarasan Ramasamy. Measuring vergence over stereoscopic video with a remote eye tracker. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*, page 97. ACM Press, 2010.

[40] Wolfgang Jaschinski and Rudolf Groner. Vergence eye movements: From basic science to clinical application: - foreword to the special issue. *Journal of Eye Movement Research*, 12(4), 2020.

[41] Maria Solé Puig, Laura Pérez Zapata, Laura Puigcerver, Neus Esperalba Iglesias, Carmen Sanchez Garcia, August Romeo, Josep Cañete Crespillo, and Hans Supèr. Attention-related eye vergence measured in children with attention deficit hyperactivity disorder. *PLOS ONE*, 10(12):e0145281, 2015.

[42] Maria Solé Puig, Laura Pérez Zapata, J. Antonio Aznar-Casanova, and Hans Supèr. A role of eye vergence in covert attention. *PLoS ONE*, 8(1):e52955, 2013.

[43] Andrew T. Duchowski, Krzysztof Krejtz, Matias Volonte, Chris J. Hughes, Marta Brescia-Zapata, and Pilar Orero. 3d gaze in virtual reality: Vergence, calibration, event detection. *Procedia Computer Science*, 207:1641–1648, 2022.

[44] Andrew T. Duchowski, Donald H. House, Jordan Gestring, Robert Congdon, Lech Świrski, Neil A. Dodgson, Krzysztof Krejtz, and Izabela Krejtz. Comparing estimated gaze depth in virtual and physical environments. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 103–110. ACM, 2014.

[45] Maria Solé Puig, August Romeo, and Hans Supèr. Vergence eye movements during figure-ground perception. *Consciousness and Cognition*, 92:103138, 2021.

[46] J. Edward Swan, Gurjot Singh, and Stephen R. Ellis. Matching and reaching depth judgments with real and augmented reality targets. *IEEE Transactions on Visualization and Computer Graphics*, 21(11):1289–1298, 2015.

[47] J. M. Foley and Richard Held. Visually directed pointing as a function of target distance, direction, and available cues. *Perception & Psychophysics*, 12(3):263–268, 1972.

[48] Sergio S. Fukusima, Jack M. Loomis, and José A. Da Silva. Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1):86–100, 1997.

[49] Adam Jones, J. Edward Swan, Gurjot Singh, and Eric Kolstad. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *2008 IEEE Virtual Reality Conference*, pages 267–268, 2008.

[50] Jonathan W. Kelly, Lucia A. Cherep, and Zachary D. Siegel. Perceived space in the HTC vive. *ACM Transactions on Applied Perception*, 15(1):1–16, 2018.

[51] Sarah H. Creem-Regehr, Jeanine K. Stefanucci, William B. Thompson, Nathan Nash, and Michael McCardell. Egocentric distance perception in the oculus rift (DK2). In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, pages 47–50. ACM, 2015.

[52] M.A. Livingston, Zhuming Ai, J.E. Swan, and H.S. Smallman. Indoor vs. outdoor depth perception for mobile augmented reality. In *2009 IEEE Virtual Reality Conference*, pages 55–62. IEEE, 2009. ISSN: 1087-8270.

[53] Stephen R. Ellis and Urs J. Bucher. Distance perception of stereoscopically presented virtual objects optically superimposed on physical objects by a head-mounted see-through display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 38(19):1300–1304, 1994.

[54] Microsoft. About hololens 2. [Online], 2023. Accessed: April 10, 2025.

[55] Microsoft. Mixed reality - comfort. [Online], 2021. Accessed: April 10, 2025.

[56] Magic Leap Inc. Learn more about magic leap 2. [Online], 2024. Accessed: April 10, 2025.

[57] Magic Leap Inc. Comfort and content placement. [Online], 2024. Accessed: April 10, 2025.

[58] Microsoft. Eye tracking on hololens 2. [Online], 2023. Accessed: April 10, 2025.

[59] Magic Leap Inc. Eye tracking. [Online], 2025. Accessed: April 10, 2025.

[60] Farzana Alam Khan, Veera Venkata Ram Murali Krishna Rao Muvva, Dennis Wu, Mohammed Safayet Arefin, Nate Phillips, and J. Edward Swan. Measuring the perceived three-dimensional location of virtual objects in optical see-through augmented reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 109–117, 2021.

[61] Farzana Alam Khan, Mohammed Safayet Arefin, Nate Phillips, and J. Edward Swan. A replication study to measure the perceived three-dimensional location of virtual objects in optical see through augmented reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 796–797, 2022.

[62] Mohammed Safayet Arefin, Nate Phillips, Alexander Plopski, Joseph L. Gabbard, and J. Edward Swan. Impact of ar display context switching and focal distance switching on human performance: Replication on an ar haploscope. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 571–572, 2020.

[63] Mohammed Safayet Arefin, Nate Phillips, Alexander Plopski, and J. Edward Swan. Effects of a distracting background and focal switching distance in an augmented reality system. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 96–99. IEEE.

[64] Wolfgang Jaschinski-Kruza. Eyestrain in vdu users: Viewing distance and the resting position of ocular muscles. *Human Factors*, 33(1):69–83, 1991. PMID: 2037310.

[65] Yasuhiro Inuzuka, Yoshimasa Osumi, and Hiroaki Shinkai. Visibility of head up display (hud) for automobiles. *Proceedings of the Human Factors Society Annual Meeting*, 35(20):1574–1578, 1991.

[66] Kenneth W. Gish and Loren K. Staplin. Human factors aspects of using head up displays in automobiles: A review of the literature, interim report. *Engineering Psychology*, 1995.

[67] James E. Cutting and Peter M. Vishton. Chapter 3 - perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth*. In William Epstein and Sheena Rogers, editors, *Perception of Space and Motion*, Handbook of Perception and Cognition, pages 69–117. Academic Press, 1995.

[68] Walter R. Miles. Ocular dominance in human adults. *The Journal of General Psychology*, 3(3):412–430, 1930. Publisher: Routledge.

[69] F. Faul, E. Erdfelder, A Buchner, and G. Lang, A. *Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses*. Behavior Research Methods, 2009.

[70] ROBERT W. BALOH, ANDREW W. SILLS, WARREN E. KUMLEY, and VICENTE HONRUBIA. Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology*, 25(11):1065–1065, 1975.

[71] Michael A. Lawrence. *ez: Easy Analysis and Visualization of Factorial Experiments*, 2016. R package version 4.4-0.

[72] Andri Signorell. *DescTools: Tools for Descriptive Statistics*, 2024. R package version 0.99.54.

[73] Daniël Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol*, 4:863, November 2013.

[74] J Martin Bland and Douglas G Altman. Statistics notes: Transformations, means, and confidence intervals. *BMJ*, 312(7038):1079, 1996.

[75] V. V. KRISHNAN, DOUGLAS SHIRACHI, and LAWRENCE STARK. Dynamic Measures of Vergence Accommodation. *Optometry and Vision Science*, 54(7), 1977.

[76] David Dunn, Okan Tursun, Hyeonseung Yu, Piotr Didyk, Karol Myszkowski, and Henry Fuchs. Stimulating the human visual system beyond real world performance in future aug-

mented reality displays. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 90–100, 2020.

[77] Mark Mon-Williams and James R. Tresilian and. Ordinal depth information from accommodation? *Ergonomics*, 43(3):391–404, 2000. PMID: 10755661.

[78] George K. Hung, Kenneth J. Ciuffreda, and Mark Rosenfield. Proximal contribution to a linear static model of accommodation and vergence. *Ophthalmic and Physiological Optics*, 16(1):31–41, 1996.

[79] Jaya Surya Bontha and Mohammed Safayet Arefin. Effort to replicate custom-made augmented reality haploscope. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 37–41, 2024.

Appendix A

License

Colorado State University LaTeX Thesis Template

by Elliott Forney – 2017

This is free and unencumbered software released into the public domain.

Anyone is free to copy, modify, publish, use, compile, sell, or distribute this software, either in source code form or as a compiled binary, for any purpose, commercial or non-commercial, and by any means.

In jurisdictions that recognize copyright laws, the author or authors of this software dedicate any and all copyright interest in the software to the public domain. We make this dedication for the benefit of the public at large and to the detriment of our heirs and successors. We intend this dedication to be an overt act of relinquishment in perpetuity of all present and future rights to this software under copyright law.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.