

DEFORMABLE BEAMSPLITTERS: ENHANCING PERCEPTION WITH  
WIDE FIELD OF VIEW, VARIFOCAAL AUGMENTED REALITY DISPLAYS

David Dunn

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in  
partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department  
of Computer Science.

Chapel Hill  
2019

Approved by:

Henry Fuchs

David Luebke

Karol Myszkowski

Montek Singh

Turner Whitted

© 2019  
David Dunn  
ALL RIGHTS RESERVED



## **ABSTRACT**

David Dunn: Deformable Beamsplitters: Enhancing Perception with Wide Field of View,  
Varifocal Augmented Reality Displays  
(Under the direction of Henry Fuchs)

An augmented reality head-mounted display with full environmental awareness could present data in new ways and provide a new type of experience, allowing seamless transitions between real life and virtual content. However, creating a light-weight, optical see-through display providing both focus support and wide field of view remains a challenge.

This dissertation describes a new dynamic optical element, the deformable beamsplitter, and its applications for wide field of view, varifocal, augmented reality displays. Deformable beamsplitters combine a traditional deformable membrane mirror and a beamsplitter into a single element, allowing reflected light to be manipulated by the deforming membrane mirror, while transmitted light remains unchanged. This research enables both single element optical design and correct focus while maintaining a wide field of view, as demonstrated by the description and analysis of two prototype hardware display systems which incorporate deformable beamsplitters. As a user changes the depth of their gaze when looking through these displays, the focus of virtual content can quickly be altered to match the real world by simply modulating air pressure in a chamber behind the deformable beamsplitter; thus ameliorating vergence–accommodation conflict.

Two user studies verify the display prototypes’ capabilities and show the potential of the display in enhancing human performance at quickly perceiving visual stimuli. This work shows that near-eye displays built with deformable beamsplitters allow for simple optical designs that enable wide field of view and comfortable viewing experiences with the potential to enhance user perception.

To Kelly, for putting up with four boys

## ACKNOWLEDGEMENTS

First off, I need to thank all of my coauthors and collaborators for guiding my research, helping build prototypes, designing user studies, and writing papers. Thanks go to Cary Tippetts for lighting a fire under a foundering project, Kent Torell for doing all the things I didn't have time for, Petr Kellnhofer for writing all the code I couldn't, Kaan Akşit for showing me what hardware research can be, Piotr Didyk for seeing the flaws in our perception, Qian Dong in carrying on for Cary, Praneeth Chakravarthula for illuminating the dark pixels, Hyesoneung Yu for gathering the far-flung literature, and Okan Tursun for the speed of his sight.

I must also thank my mentors Joohwan Kim, Ward Lopes, Josef Spjut, and Quinn Smithwick for guiding my research, providing useful feedback, and generally being great guys.

The institutions of NVIDIA Research and Disney Research deserve thanks for providing me such wonderful opportunities to meet new researchers and collaborate on interesting projects through several internships. I also enjoyed getting paid and being able to support my family over the summers and through some of the school years. Additionally, I must thank the National Science Foundation and the Fraunhofer and Max Planck cooperation program who partially supported this work.

I have the greatest doctoral committee; each of them has provided me great feedback and solid guidance. Thank you David Luebke, Karol Myszkowski, Montek Singh, and Turner Whitted. Of course I can't go without mentioning my adviser, Henry Fuchs, who not only believed in me enough to take me on as a student, but provided sound direction and support.

Jim Mahaney deserves a very large thank you for obtaining all the materials, running all the miles, spouting all the wise-cracks, eating all the chicken wings, and assisting me in constructing my often foolish ideas.

Thank you to my Creator and Savior for giving me all that I have and making me all that I am.

Lastly, I must thank my family. Thanks go to Kelly, my wife, for putting up with all my nonsense. To my parents, Ann and Gary, thank you for raising me correctly and supporting me. To Adam, Daniel, and Joseph, thank you for always surprising me and keeping life exciting; I am proud to be your father.

## PREFACE

I have always been interested in making magical experiences for others, whether it was by watching the animated characters I worked on come to life or in using a tool I had written, I enjoyed making an impact in peoples lives. Once I decided to return to school to pursue a graduate education, I knew I wanted to continue making these magical experiences.

Before enrolling in my first courses, I had some ideas on how augmented reality could bring those same animated characters to life. It should be possible to allow them to interact with and inhabit the same space as the user, if only there were a device capable of it. In my innocence, I thought making such a device would be straight forward, and even drew out designs for a few prototypes. Little did I know, both physics and physiology were against me and that I needed much more knowledge before I would be able to create a successful design.

Since then, I have learned a great deal both about the topic of near-eye displays and about being a graduate student. The work presented in this dissertation comes from three of the papers I have published over the course of my time here at UNC-CH. Those three papers are *Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors* (Dunn et al., 2017), *Mitigating Vergence-Accommodation Conflict for Near-Eye Displays via Deformable Beamsplitters* (Dunn et al., 2018), and *Required Accuracy of Gaze Tracking for Varifocal Displays* (Dunn, 2019). While there have been many failures along the road, there has also been much rejoicing for successes, and I look forward to many more successes and failures in my future.

While I haven't yet achieved my dream of providing real-world animated character interactivity, I believe the technology we have worked on is providing one more stone in the foundation of what will hopefully come soon.

## TABLE OF CONTENTS

LIST OF TABLES .....	xiii
LIST OF FIGURES .....	xiv
LIST OF ABBREVIATIONS .....	xvii
CHAPTER 1: INTRODUCTION .....	1
1.1 Augmented Reality .....	2
1.2 Enabling Technologies .....	2
1.3 Near-Eye Displays for Augmented Reality .....	3
1.4 Thesis Statement .....	5
1.5 Contributions .....	6
1.6 Structure .....	6
CHAPTER 2: BACKGROUND .....	7
2.1 Monocular Human Visual System .....	8
2.1.1 Angular Resolution .....	8
2.1.2 Field of View .....	10
2.1.3 Temporal Resolution .....	10
2.1.4 Eye Movements .....	12
2.1.5 Focus and Blur .....	12
2.1.6 Wavelength .....	13
2.1.7 Dynamic Range .....	14
2.1.8 Polarization .....	14
2.2 Binocular Human Visual System .....	15

2.2.1	Vergence–Accommodation Conflict .....	15
2.2.2	Panum’s Fusion Area .....	16
2.2.3	Depth of Field / Depth of Focus .....	16
2.2.4	Zone of Clear Single Binocular Vision .....	18
2.2.5	Zone of Comfort .....	18
2.2.6	Accuracy of Vergence and Accommodation .....	18
2.2.7	Speed of Vergence and Accommodation .....	19
2.3	Considerations in Developing Near-Eye Displays .....	20
2.3.1	Étendue or Throughput .....	23
2.3.2	Focal Range Trade-off .....	24
2.3.3	Real-world Transmission .....	25
2.3.4	Visual Quality .....	25
2.4	Prior Work .....	25
2.4.1	Maxwellian View Displays .....	27
2.4.2	Virtual Retinal Displays .....	31
2.4.3	Multifocal Displays .....	32
2.4.4	Varifocal Displays .....	33
2.4.5	Light Field Displays .....	34
2.4.6	Holographic Displays .....	35
2.5	Discussion .....	36
CHAPTER 3: DEFORMABLE BEAMSPLITTERS – A NEW OPTICAL ELEMENT .....		37
3.1	Previous Work in Deformable Membrane Mirrors .....	37
3.2	Developing Our Early Prototypes .....	39
3.2.1	Membrane .....	40
3.2.2	Housing .....	44
3.2.3	Assembly .....	46

3.3	Properties of Deformable Beamsplitters .....	48
3.3.1	Optical Properties .....	48
3.3.2	Deformation Properties .....	49
3.4	Customization for Near-Eye Displays .....	53
3.4.1	Ray tracing model .....	54
3.4.2	Design space .....	56
3.5	Discussion .....	60
CHAPTER 4: NEAR-EYE DISPLAY PROTOTYPES.....		61
4.1	First Prototype Display .....	62
4.1.1	Image Generation .....	63
4.1.2	Optics .....	64
4.1.3	Membrane Actuation.....	65
4.1.4	Membrane Feedback .....	66
4.1.5	User Focus Detection .....	66
4.1.6	Limitations .....	67
4.2	Second Prototype Display.....	70
4.2.1	Image Generation .....	72
4.2.2	Optics .....	73
4.2.3	Membrane Actuation.....	75
4.2.4	Membrane Feedback .....	76
4.2.5	User Focus Detection .....	76
4.2.6	Optical Quality Analysis .....	77
4.3	Discussion .....	82
CHAPTER 5: USER FOCUS DETECTION .....		83
5.1	Measuring Focal State .....	83
5.2	Determining Error .....	84



5.2.1	Calculating Distance of Fixation .....	84
5.2.2	Error Assumptions .....	87
5.2.3	Differences Between virtual reality (VR) and augmented reality (AR) .....	88
5.3	Evaluation of Near-Eye Gaze Trackers .....	89
5.3.1	Intrusive Eye Gaze Trackers .....	89
5.3.2	Non-intrusive Eye Gaze Trackers .....	90
5.3.3	Remote Eye Gaze Trackers .....	91
5.4	Discussion .....	92
CHAPTER 6: PERCEPTUAL VALIDATION AND ENHANCEMENT .....		93
6.1	Monocular Acuity Study .....	93
6.1.1	Hypothesis .....	93
6.1.2	Experiment Configuration .....	93
6.1.3	Stimuli .....	94
6.1.4	Participants .....	95
6.1.5	Procedure .....	95
6.1.6	Results .....	96
6.2	Over-driving Focus .....	98
6.2.1	Hypothesis .....	99
6.2.2	Experimental Configuration .....	99
6.2.3	Stimuli .....	99
6.2.4	Participants .....	100
6.2.5	Procedure .....	101
6.2.6	Results .....	102
6.2.7	Discussion .....	103
6.3	Over-driving Focus and Vergence .....	103
6.3.1	Hypothesis .....	103

6.3.2	Experimental Configuration .....	104
6.3.3	Stimuli .....	104
6.3.4	Participants .....	106
6.3.5	Procedure .....	106
6.3.6	Results .....	108
6.3.7	Discussion.....	109
6.4	Over-driving Focus and Vergence with Saccade .....	110
6.4.1	Hypothesis .....	111
6.4.2	Experimental Configuration .....	111
6.4.3	Stimuli .....	112
6.4.4	Participants .....	112
6.4.5	Procedure .....	112
6.4.6	Results .....	112
6.4.7	Discussion.....	113
CHAPTER 7:	SUMMARY AND CONCLUSION .....	115
7.1	Future Work .....	116
7.2	Conclusion.....	118
REFERENCES	.....	119

## LIST OF TABLES

Table 2.1 – Comparison of Near-Eye Displays with Accommodative Cues Part 1 .....	28
Table 2.2 – Comparison of Near-Eye Displays with Accommodative Cues Part 2 .....	29
Table 4.1 – Prototype 2 luminance .....	81
Table 4.2 – System comparison for two prototype displays .....	81
Table 4.3 – Parameter comparison of prototypes .....	82
Table 5.1 – Required Gaze Tracking Accuracy in Central Field .....	88
Table 5.2 – Reported Gaze Tracking Accuracy.....	91

## LIST OF FIGURES

Figure 1.1 – Scenario depicting need for focus support in augmented reality .....	4
Figure 2.1 – Anatomy of the human eye.....	9
Figure 2.2 – Campbell-Robson chart showing spatial contrast sensitivity function of the human eye.....	9
Figure 2.3 – Density of photoreceptors along the eccentricity of the human eye .....	10
Figure 2.4 – Average Monocular Human Field of View .....	11
Figure 2.5 – Average Binocular Human Field of View .....	11
Figure 2.6 – Eye sensitivity across spectrum .....	13
Figure 2.7 – Eye dynamic range .....	14
Figure 2.8 – Zones of binocular fusion and single vision .....	17
Figure 2.9 – Classes of near eye displays capable of large focal range.....	30
Figure 3.1 – Example deformable membrane mirror display from 1982 .....	38
Figure 3.2 – Sketch of deformable beamsplitter display optical layout for a single eye .....	39
Figure 3.3 – Poor image quality from PVC film .....	41
Figure 3.4 – Deformable beamsplitter with PVC film and PVC housing.....	42
Figure 3.5 – Difference in reflection intensity after membrane metalization .....	42
Figure 3.6 – Membrane surface quality .....	43
Figure 3.7 – Deformable beamsplitter using PDMS membrane .....	45
Figure 3.8 – Deformable beamsplitter housing produced by FDM printing .....	46
Figure 3.9 – Deformable beamsplitter housing produced by PolyJet printing .....	46
Figure 3.10 –Exploded diagram of deformable beamsplitter housing .....	47
Figure 3.11 –Transmission and reflection characteristics of membrane .....	48

Figure 3.12 –Stress-strain relationship for PDMS membrane.....	50
Figure 3.13 –Shape of the membrane as it deforms .....	51
Figure 3.14 –Stress test of membrane.....	52
Figure 3.15 –Computed field of view for optical configurations .....	57
Figure 3.16 –Computed properties for different eye reliefs at specified focal depths .....	58
Figure 3.17 –Computed properties for different display distances at specified focal depths....	59
 Figure 4.1 – System overview .....	 61
Figure 4.2 – First prototype field of view (FOV) .....	63
Figure 4.3 – First prototype system overview .....	64
Figure 4.4 – First prototype display .....	65
Figure 4.5 – First prototype vacuum system .....	66
Figure 4.6 – First prototype control electronics .....	67
Figure 4.7 – First prototype point spread photograph .....	69
Figure 4.8 – First prototype displaying teapots in hand at several depths .....	71
Figure 4.9 – View through second display prototype .....	71
Figure 4.10 –Second prototype display .....	72
Figure 4.11 –Second prototype system overview .....	73
Figure 4.12 –Close up view of rendered blur .....	74
Figure 4.13 –Membrane control circuit in second prototype.....	76
Figure 4.14 –Second prototype vacuum system .....	77
Figure 4.15 –Field of view of both prototype displays .....	78
Figure 4.16 –Second prototype modulation transfer function .....	80
 Figure 5.1 – Calculating the distance of fixation using trigonometric methods .....	 85
Figure 5.2 – Diopters of error due to gaze tracking error .....	87

Figure 6.1 – Monocular acuity study experiment configuration .....	94
Figure 6.2 – Virtual and real co-located example stimuli .....	94
Figure 6.3 – Virtual and real co-located task performance .....	96
Figure 6.4 – Over-driving focus experiment configuration .....	98
Figure 6.5 – View of monocular over-driving focus experiment display .....	99
Figure 6.6 – Over-driving focus experimental procedure .....	100
Figure 6.7 – Over-driving focus experiment results .....	102
Figure 6.8 – Over-driving accommodation and vergence configuration .....	104
Figure 6.9 – View of binocular vergence and accommodation over-driving experiment display .....	105
Figure 6.10 –Over-driving accommodation and vergence experimental procedure .....	107
Figure 6.11 –Over-driving focus and vergence experiment latency .....	109
Figure 6.12 –Over-driving focus and vergence experiment results .....	110
Figure 6.13 –Over-driving accommodation and vergence with saccade experimental procedure .....	111
Figure 6.14 –Over-driving focus and vergence with saccade experiment latency .....	113
Figure 6.15 –Over-driving focus and vergence with saccade experiment results .....	114
Figure 7.1 – Scenario depicting correct focus support in augmented reality .....	117

## LIST OF ABBREVIATIONS

2D	Two-Dimensional
3D	Three-Dimensional
AR	Augmented Reality
CFF	Critical Flicker Fusion
CIE	International Commission on Illumination
cpd	Cycles per Degree
DC	Direct Current
D	Diopter
DMD	Digital Micromirror Display
DMMD	Deformable Mirror Membrane Device
DOF	Depth of Field or Depth of Focus
EOG	Electrooculography
FDM	Fused Deposition Modeling
FOV	Field of View
fps	Frames per Second
FWHM	Full Width at Half Maximum
HMD	Head-Mounted Display
HVS	Human Visual System
IPD	Interpupillary Distance
IR	Infrared Radiation
IROG	Infrared Oculography
LCD	Liquid Crystal Display
LED	Light-Emitting Diode
MEMS	Micro-Electromechanical System
ML	Machine Learning
MTF	Modulation Transfer Function

NED	Near-Eye Display
OST	Optical See-through
PC	Personal Computer
PDMS	Polydimethylsiloxane
PET	Polyethylene Terephthalate
PFA	Panum's Fusion Area
PID	Proportional Integral Derivative
ppi	Pixels per Inch
PVC	Polyvinyl Chloride
REGT	Remote Eye Gaze Trackers
rpm	Revolutions per Minute
Si	Silicon
SLA	Stereolithography Apparatus
SLM	Spatial Light Modulator
SSC	Scleral Search Coils
TFT	Thin Film Transistor
UNC-CH	University of North Carolina at Chapel Hill
USB	Universal Serial Bus
VAC	Vergence–Accommodation Conflict
VOG	Video Oculography
VR	Virtual Reality
VST	Video See-through
ZCSBV	Zone of Clear Single Binocular Vision
ZOC	Zone of Comfort



## CHAPTER 1: INTRODUCTION

With the advent of the modern computing era, as computers have developed from mechanical machines, through mainframes and personal computers, to our modern pocket-sized and internet-connected devices, there has been an accompanying progression in the methods of human-computer interaction. In the early days of modern computing, information transfer was limited to computer-prioritized batch interfaces such as paper tape, punch cards, and line printers. Even command-line interfaces with keyboards and text displays, which greatly increased the ease of communication between human and machine, mostly favored the machine by using unintuitive commands. As more computational power became available, an increasing amount could be spent on improving the human interfaces, leading to graphical user interfaces with mice, stylus, and graphical displays. In the last decade and a half, user-prioritized mobile touch interfaces with hand-held touch displays and conversational interfaces with voice and speaker have become commonplace.

More computationally demanding interfaces have been around for decades but are not yet mainstream. With the rise of consumer-level virtual reality (VR) hardware, the next interfaces to move from the exhibition hall to the living room are spatial, motion-tracked interfaces, which use both motion controllers (hand and head-mounted) and 3D immersive displays. These new interfaces, which allow information to be displayed and interaction to occur in a 3D environment with spatially registered elements, provide a larger set of potential interaction techniques. We spend our lives in a spatial world interacting with physical objects. These new spatial motion-tracked interfaces can potentially provide natural interactions akin to our daily lives, allowing for a more natural transfer of information. If we seek to combine real-world elements into these interactions, then augmented reality (AR) affords us that possibility.

## **1.1 Augmented Reality**

Azuma (1997) defined AR as requiring three elements. It must (1) combine real and virtual content, (2) be interactive in real time, and (3) be spatially registered in three dimensions.

AR has the potential to drastically change our daily lives. An interface with full environmental integration could present new data and experiences allowing seamless transitions between real and virtual. Commercial AR is expected to open a multitude of possibilities by bridging the gap between computer graphics and human vision – creating virtual content that is indistinguishable from the real world.

In the medical field, several works have been presented illustrating how AR may enhance surgical procedures and rehabilitation (Ilie et al., 2004). Industrial applications include simplifying complex factory part assembly and training (Henderson and Feiner, 2011). In the commercial sector, explorations in a user seeing their reflection wearing a virtual outfit or being directed to a restaurant in a unfamiliar locale have been described (Scholz and Smith, 2016; Narzt et al., 2006). AR can enable better communication via more natural face-to-face telepresence (Wang et al., 2014). Many look toward entertainment as one of the largest markets for AR with descriptions of children playing at home with their favorite virtual characters (Vera et al., 2011), or experiencing cinema in the real world around us (MacIntyre et al., 2001). In short, by altering the very reality we experience, AR is a transformative technology.

## **1.2 Enabling Technologies**

With the wide potential of AR, it is no wonder that many decades of research have sought to create devices which can create those experiences. Many technologies are required to work in concert in order to create an effective AR interface, including user tracking, device tracking, environmental mapping, reconstruction, and display. The best devices seamlessly integrate each of these systems into a natural user experience. The display acts as the visual bridge between

computer and human, and while it is dependent upon the other technologies, it will be the sole topic of this work.

Several different categories of displays exist with different capabilities and hardware requirements. In broad terms they can be divided into spatial AR (projector-based displays), hand-held displays, video see-through (VST) near-eye displays (NEDs), and optical see-through (OST) NEDs. For spatial AR, the light engine is a projector not co-located with the user, which illuminates real-world objects. Common examples are sandtables and robotic avatars (Lincoln et al., 2011). Hand-held displays are the most common form of augmented reality available today. With phone or tablet applications such as Ikea Place<sup>1</sup> and Pokémon Go<sup>2</sup>, millions of users can experience augmented reality with the devices they carry around every day (Scholz and Smith, 2016). With the recent resurgence of commercial VR devices, a simple way to extend them for AR applications is to provide a VST component from a head-mounted camera. Commercial devices such as the Vive Pro<sup>3</sup> and Windows MR<sup>4</sup> headsets provide this additional functionality. The advantage of OST NEDs is in providing unobstructed views of the real world in a non-encumbering, hands-free manner. However, the difficulties in designing a device capable of combining real-world light with the light of virtual imagery in an acceptable manner will be described in section 2.3.

### **1.3 Near-Eye Displays for Augmented Reality**

Current available commercial OST AR NEDs like Microsoft HoloLens<sup>5</sup> and Magic Leap One<sup>6</sup> are successful in overlaying high quality virtual images onto the real world, but have limitations in terms of available field of view (FOV) and resolution. The HoloLens presents stereoscopic imagery without correct focal cues, while the Magic Leap One has begun to address the lack of

---

<sup>1</sup> <https://highlights.ikea.com/2017/ikea-place/>

<sup>2</sup> <https://www.pokemongo.com/en-us/>

<sup>3</sup> <https://www.vive.com/us/product/vive-pro/>

<sup>4</sup> <https://www.microsoft.com/en-us/windows/windows-mixed-reality>

<sup>5</sup> <https://www.microsoft.com/en-us/hololens>

<sup>6</sup> <https://www.magicleap.com/magic-leap-one>



Figure 1.1: A scenario depicting need for focus support in augmented reality. Both the restaurant and the pamphlet annotations are focused at a single, far, depth. When a viewer focuses on the far buildings (top), all annotations are clear and sharp, however when a viewer focuses on the near pamphlet (bottom), the annotations, even those indicating the pamphlet, are blurred.

focal cues by providing a display with two focal planes. While an improvement, two focal planes are still far from what the real world provides, and the device has a similarly limited FOV. Figure 1.1 depicts a scenario that illustrates the need for focus support in AR NEDs.

The need for a wide FOV has been demonstrated by many research studies (Arthur, 2000), but the following simple logic exercise presents a compelling argument. Visual augmentations can have many purposes, among which are calling attention to and providing additional information about real world objects. If a display is capable of only presenting augmentations in a small field — even if the real world is much more widely visible — it becomes difficult to call attention to objects outside the current FOV. It would require the user to turn their head so that an object falls in the view frustum of the display for any attention-calling and detail-providing augmentation to appear, and much augmented information may go unviewed. While additional interface cues may be provided indicating off-screen content (Roberts et al., 2013), in a densely annotated environment these cues just become additional noise and are insufficient to the task. A wide FOV enables attention-calling augmentations in the periphery of the visual field to perform their task as designed.

This dissertation presents a solution to both of these problems: deformable beamsplitters, a new type of dynamic optical element. Traditional beamsplitters, or half-silvered mirrors, have been used in many AR applications from NEDs to Pepper’s ghost (Groth, 2007). Deformable beamsplitters combine a traditional deformable membrane mirror and a beamsplitter into a single element, allowing reflected light to be manipulated by the deforming membrane mirror, while transmitted light remains unchanged. Displays using deformable beamsplitters are more practical and inexpensive than comparable focus providing NEDs.

## **1.4 Thesis Statement**

Through the use of deformable beamsplitters for see-through near-eye displays, it is possible to create single optical-element, varifocal displays with increased field of view which are

better able than previous designs to integrate virtual imagery with the real world in measurable, perceptually advantageous ways.

## **1.5 Contributions**

The primary contribution of this dissertation is the description and characterization of a new dynamic optical element, the deformable beamsplitter, which can adjust the focus of reflected light while not affecting transmitted light. Its application to NEDs is discussed and two prototypes which provide wide FOV and extended focal range in a small, light-weight formfactor. The potential for driving these NEDs through the use of eye tracking is examined and determined to not be feasible with currently available eye tracking solutions. A user study validates the capability of our prototypes in improving accuracy during mixed reality tasks. Additionally, potential enhancements to human perception enabled by the new optical element are explored.

## **1.6 Structure**

Chapter 2 provides background on the human visual system (HVS), considerations in designing NEDs, and a review of previous work in the area. Chapter 3 describes deformable beamsplitters, their fabrication, and implications on NED design. Chapter 4 reviews the two prototype deformable beamsplitter displays which were first presented in Dunn et al. (2017) and Dunn et al. (2018). Chapter 5 describes the method for determining the focal state of the user as described in Dunn (2019). Chapter 6 outlines a series of user studies that validate the display capabilities and explore potential HVS enhancements which deformable beamsplitters can provide. Chapter 7 provides a summary and conclusion.

## CHAPTER 2: BACKGROUND

In his assessment of the state of VR in 1999, Fred Brooks paraphrased a lecture from 1965 given by Ivan Sutherland (Sutherland, 1965):

Don't think of that thing as a *screen*, think of it as a *window*, a window through which one looks into a *virtual world*. The challenge to computer graphics is to make that virtual world look real, sound real, move and respond to interaction in real time, and even feel real. (Brooks, 1999)

If we are to take Sutherland literally, then our goal for making the virtual world look real is to calculate and recreate every property for all light that would pass from a virtual world through our display window and into the pupils of the eyes. The light passing through a classical window can be deconstructed into a set of rays of light. A single ray of light  $r$  can be represented as

$$r = (x, y, z, t, \theta, \phi, \lambda, \psi, \chi) \quad (2.1)$$

where  $x$ ,  $y$ , and  $z$  are the coordinates of the emission point,  $t$  is the emission time,  $\theta$  and  $\phi$  are the emission angle,  $\lambda$  is the wavelength of the light, and  $\psi$  and  $\chi$  are the polarization state.

Luckily, recreating all of these properties is not required. The HVS is not sensitive to the sum properties of light for the entire set of rays, as can be seen by examining a simple case. While in certain conditions, the human eye can detect the polarization state of light,  $\psi$  and  $\chi$ , generally we are not sensitive to it (Temple et al., 2015). Thus two of the terms can be eliminated almost completely. Just as polarization is virtually eliminated, our eyes have a finite angular resolution and contrast sensitivity, among other things, which if taken into account, will greatly reduce the complex light function required for recreating the light passing through our virtual window. In this section I will discuss the aspects of the monocular and binocular HVS which are relevant to

NEDs (sections 2.1 and 2.2), I will review considerations required in designing NEDs (section 2.3), and I will review previous related display works (section 2.4).

## **2.1 Monocular Human Visual System**

The human eye is a complex system for collecting, focusing, and sensing a narrow band of the electro-magnetic spectrum. The clear cornea and aqueous humor act as the first optical surface collecting the light through the pupil, which acts as an aperture stop, and relaying it to the crystalline lens. The crystalline lens has an adjustable curvature which adds additional focusing power based on the tension of the ciliary muscles. The light then passes through the vitreous humor and hits the retina, which contains a layer of two types of sensor cells that are sensitive to light. The retina has several regions which are defined by different characteristics in the distribution of the two types of sensor cells. A region in the center of the field of vision, called the *fovea*, contains the highest density of retinal cones and is the area of highest visual acuity. Two additional features visible from outside the eye are the iris, which controls the size of the pupil, and the sclera, an opaque white protective layer of the eye. An illustration detailing the basic parts of the eye can be seen in Figure 2.1.

### **2.1.1 Angular Resolution**

Human visual acuity is limited by many factors including sensor cell distribution, pupil diffraction, lens aberrations, and contrast of stimulus. An illustration of the spatial contrast sensitivity function can be seen in Figure 2.2. The consensus of studies explained in Weymouth (1958) show that the maximum angular resolution in the central fovea for human with 20/20 acuity to be about  $0^{\circ}30''$  of arc or 60 cycles per degree (cpd).

Compared to the foveal region, other regions have much less visual acuity as the sensor cell distribution changes. This is consistent around the eye as a function of eccentricity from the center visual field. This can be easily described by plotting the density of cones and rods in a cross-section which intersects the fovea as can be seen in Figure 2.3.



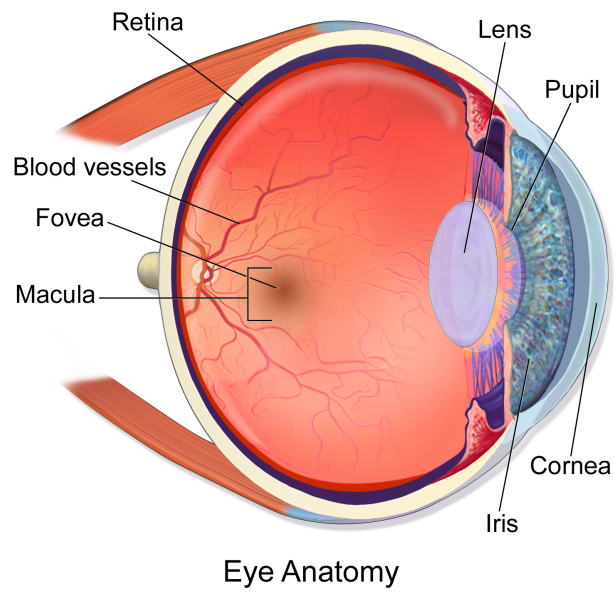


Figure 2.1: Anatomy of the human eye. Used with permission (staff, 2014).

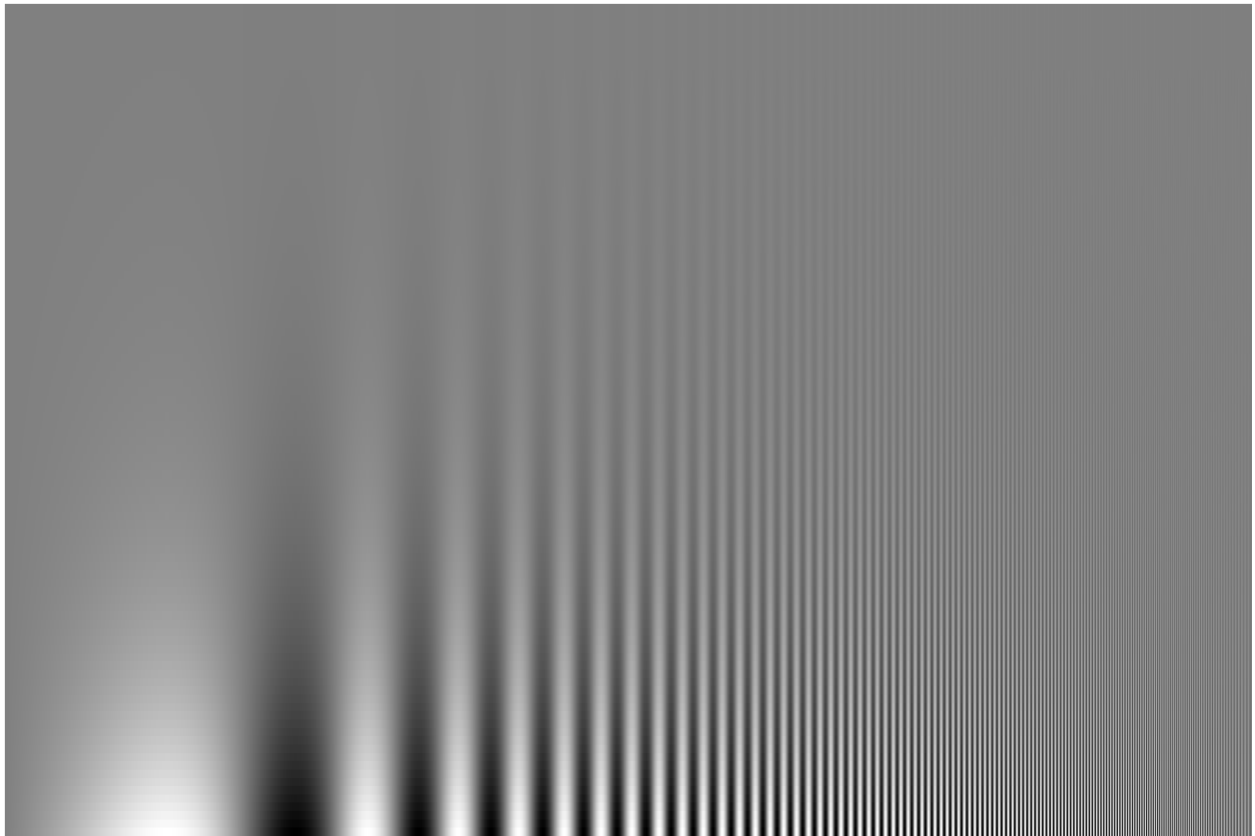


Figure 2.2: Campbell-Robson chart showing spatial contrast sensitivity function of the human eye. Note the shape of the threshold between perceived solid gray and the grating pattern.

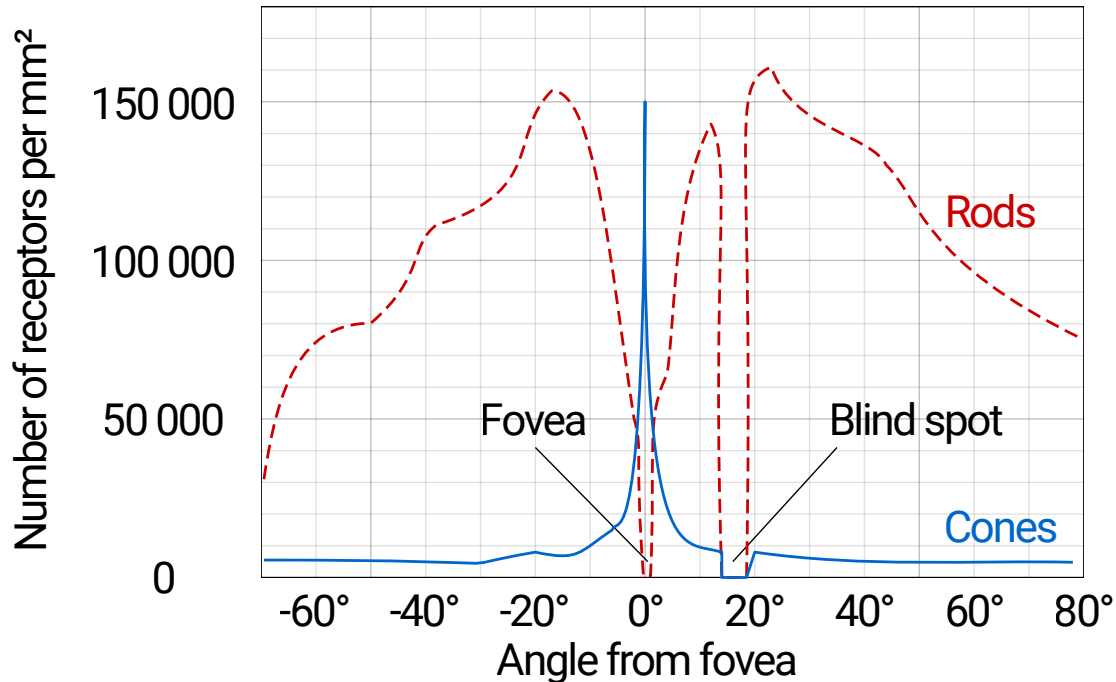


Figure 2.3: Density of rod (dotted line) and cone (solid line) photoreceptors along a line passing through the fovea and the blind spot of a human eye vs the angle measured from the fovea. Used with permission (Cmglee, 2013).

### 2.1.2 Field of View

A typical human monocular visual field extends 60° nasally limited by the nose, 60° superiorly limited by the eyebrow, 100° temporally limited by the pupil and 70° inferiorly limited by the cheek (Savino and Danesh-Meyer, 2012). This means a total monocular FOV of 160° horizontal and 130° vertical, but a polar plot more fully expresses the average FOV as seen in Figure 2.4 and Figure 2.5. When eye motion is considered, an additional 50° may become visible as the eye rotates toward the temple (Howard and Rogers, 2008).

### 2.1.3 Temporal Resolution

The eye accumulates light continuously with some amount of decay. This decay is known as *persistance of vision*, and can be tested by altering the frequency of a flashing stimulus. At low frequencies, the blinking is apparent, but above a threshold, the blinking is perceived as a solid illuminated state. This threshold is referred to as critical flicker fusion (CFF) and can vary

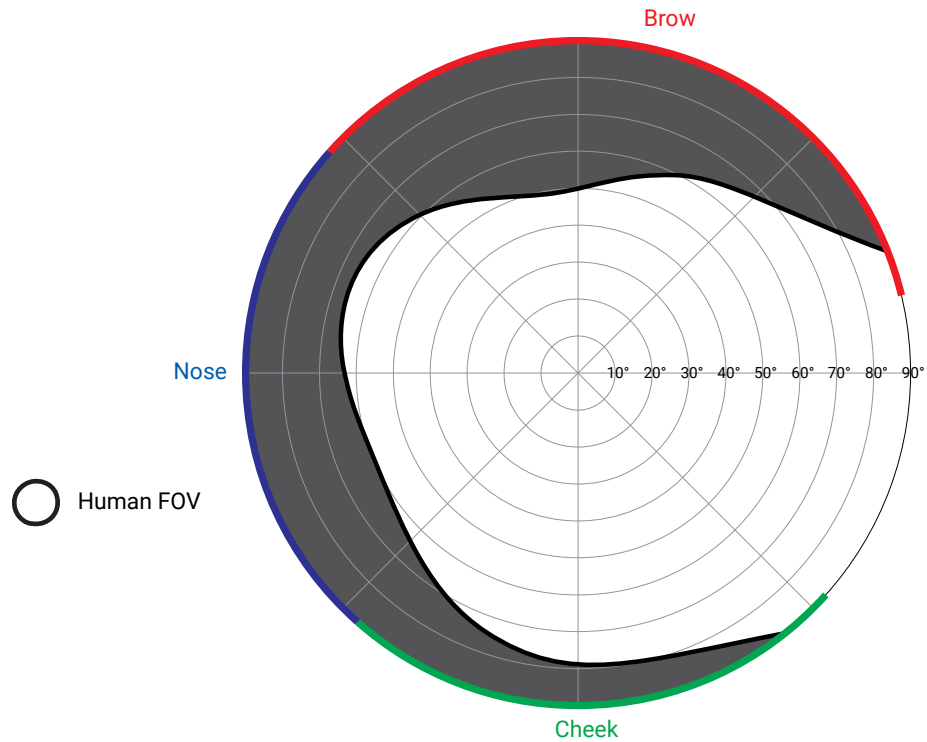


Figure 2.4: Field of view for the average human right eye, as measured from directly in front of the head. Views are blocked by the brow above, nose to the left, and cheek below (Fuchs, 1899).

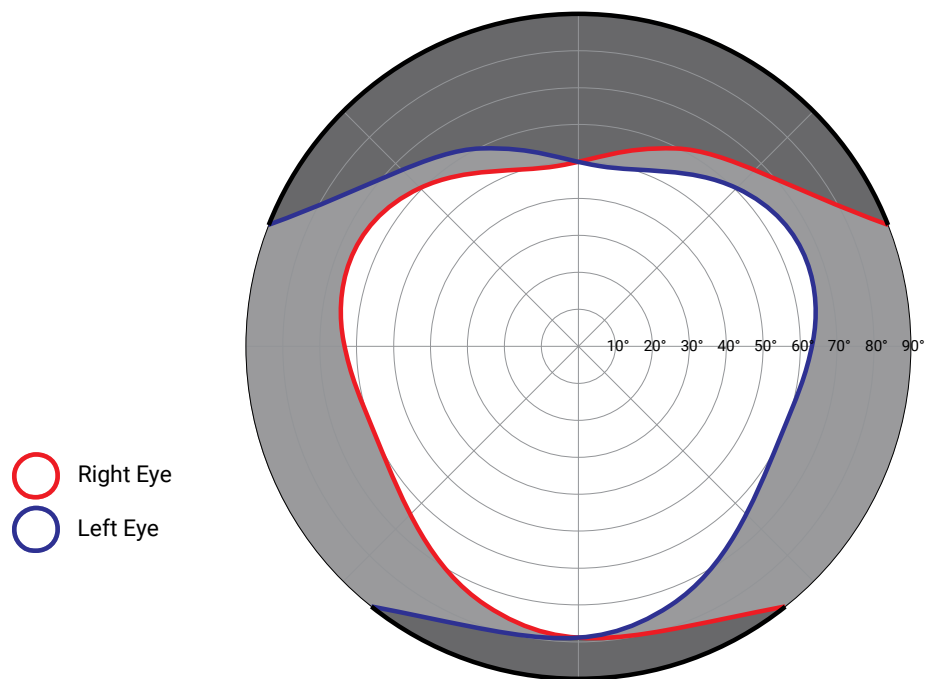


Figure 2.5: Field of view for the average human showing the stereoscopic region (white) and the monoscopic regions (light gray) (Fuchs, 1899).

based on intensity or visual solid angle of the stimulus, and the visual field eccentricity or state of fatigue in the user. The standard range for CFF is 50 Hz to 90 Hz (Landis, 1953; Farrell et al., 1987; Liu et al., 2014).

#### **2.1.4 Eye Movements**

There are four major kinds of eye movement: saccades, smooth pursuit, vergence, and vestibulo-ocular movements. *Saccades* encompass both the large motions of looking around a room and the small motions of reading a book, and both eyes always rotate in the same direction. They can be voluntarily initiated, but occur reflexively, lasting between 15 ms and 100 ms, and incur a strongly suppressed visibility, known as *saccadic suppression* or *saccadic omission*, during and shortly after the motion. *Smooth pursuit* is a tracking mechanism for following a moving stimulus and maintaining its location on the fovea. Such tracking can successfully be performed up to a certain velocity of moving objects, typically less than  $80^\circ/\text{s}$ , and in most people requires a moving stimulus. *Vergence* movements align the two eyes as the fixation point changes in depth, and are characterized by the two eyes moving in opposite directions. Their speed and characteristics are described further in section 2.2. The last classification of eye movements is *vestibulo-ocular*, which stabilize the image on the retina as head motion occurs. This compensatory reflex allows for stable percepts by offsetting the head motion with eye motion (Purves et al., 2001).

It should be noted that each of these motions has an effect on the temporal resolution of the eye. For example, when motion is involved studies have shown that flicker is perceived at frequencies higher than the CFF (Watson, 2013).

#### **2.1.5 Focus and Blur**

Human eye focal range varies by subject and changes with age. The mean focal range of a 10 year old is 13.4 D(diopter), while at 55 years a mean of 1.3 D is reported by Duane (1922). When objects are not in focus, a blurred image is perceived.

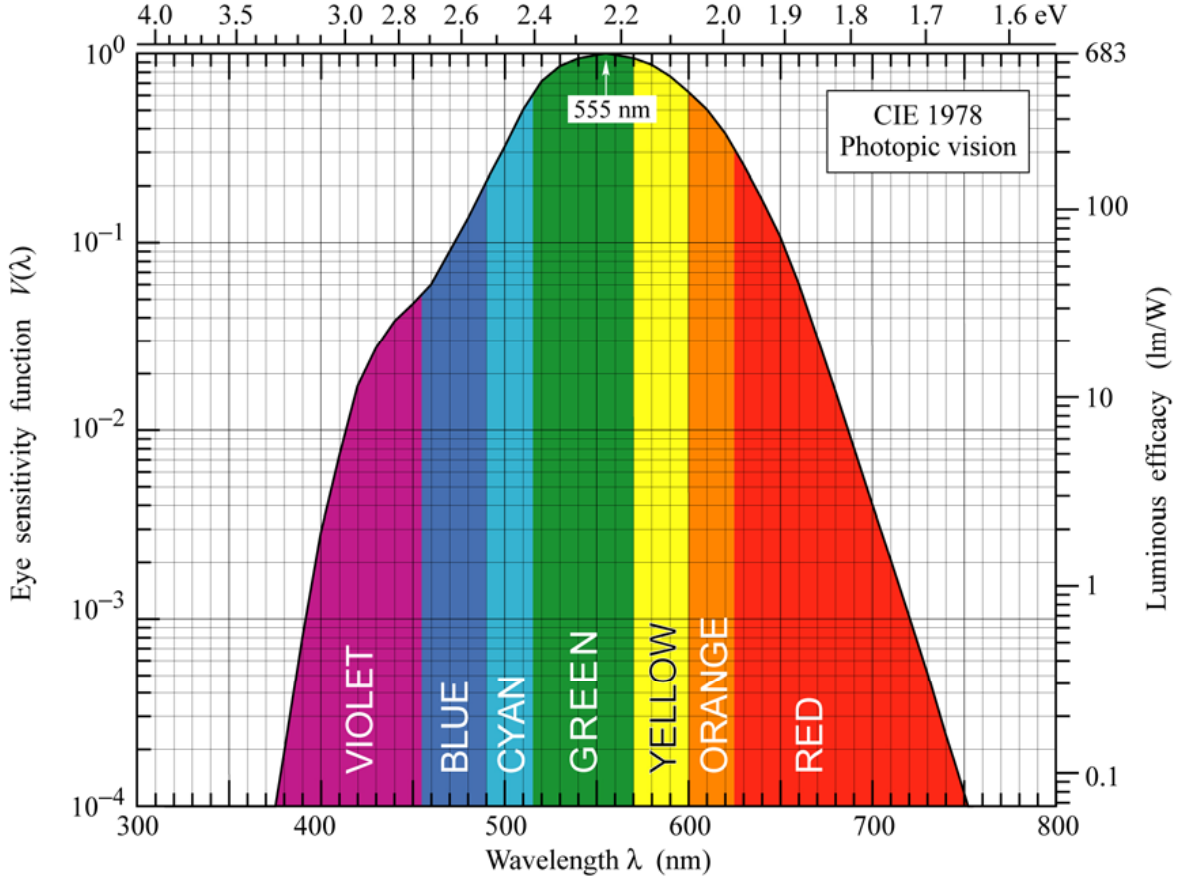


Figure 2.6: Eye sensitivity function  $V(\lambda)$  and luminous efficacy measured in lumens per watt of optical power across the light spectrum.  $V(\lambda)$  is maximum at 555 nm. From Schubert (2003)

Blur is also a very important perceptual depth cue (Zannoli et al., 2016); an accommodation-supporting, dynamic-focus display incapable of displaying multiple depths simultaneously needs to render appropriate synthetic blur on virtual objects that are at focal distances different than the current displayed depth. Recent studies show that computational blur that takes into account the chromatic aberrations of eye, dubbed Chroma-blur (Cholewiak et al., 2017), can improve perceived quality of synthetic imagery.

### 2.1.6 Wavelength

Three types of cone cells are sensitive to three different bands of the electro-magnetic spectrum, which correspond to red, green and blue. The overall sensitivity function at different wave-

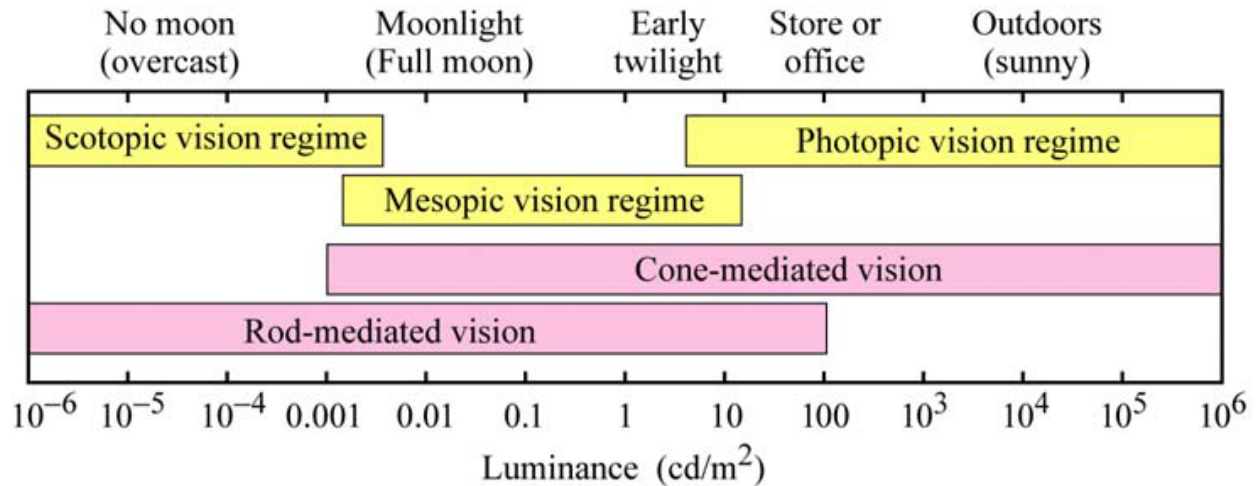


Figure 2.7: Approximate ranges of vision regimes and receptor regimes across the dynamic range of the human eye. From Schubert (2003).

lengths is described by the CIE in 1978 and can be seen in figure 2.6. The function has a maximum at 555 nm in the green range, and falls off to the red and violet (Schubert, 2003).

### 2.1.7 Dynamic Range

Due to the different sensitivities of the cone and rod cells, the human eye has a wide range of illumination sensitivity which can be broken down into 3 ranges. *Photopic* vision occurs at ranges of high ambient illumination and is mediated by the cones. *Scotopic* vision happens at low ambient illumination ranges when vision is mediated by the rods, a primary characteristic of which is loss of color sensitivity. *Mesopic* vision occurs at the illuminance levels between the photopic and scotopic ranges when vision is mediated by both cones and rods (Schubert, 2003).

### 2.1.8 Polarization

Haidinger (1844) described a visual phenomenon, now known as *Haidinger's brushes*, where humans can directly detect the orientation of polarized light. When viewing a polarized light field with no spatial variation in color or intensity, faint blue and yellow bow-tie like shapes appear. The effect vanishes within 5 s but may be maintained by moving the eye. Unlike other vision, sensitivity to polarization is mediated by dichroic carotenoids in the macula (Temple et al., 2015).

Polarization, having minimal impact on human perception, is generally ignored with respect to presenting it correctly in displays. It is often used by the display itself as a means of controlling and manipulating the light transfer and conditioning through the optical system. To the best of my knowledge, no studies have explored the effects of presenting incorrect polarization cues to a viewer in virtual or augmented environments.

## **2.2 Binocular Human Visual System**

While the visual responses in the monocular eye are easily linked to physiological elements, the mechanism for combining the images from two eyes into a single percept, referred to as *binocular fusion*, is more elusive. Many mechanisms, including the vergence and accommodation of the eyes, come into play when fusing the two images. When these mechanisms fail, the possible outcomes are referred to as *binocular rivalry*. Rivalry encompasses both *suppression*, where the brain suppresses the view from one of the eyes either spatially or time varying, and *diplopia*, where both views are visible simultaneously — commonly called double vision. For a comfortable AR experience, an AR display needs to provide appropriate perceptual cues so that proper vision may be maintained.

### **2.2.1 Vergence–Accommodation Conflict**

The vergence and accommodation of human vision are neurally coupled (Fincham and Walton, 1957), and presenting unmatching depth and vergence cues will force a decoupling leading to an effect commonly known as *vergence–accommodation conflict* (VAC). Many recent studies show that VAC is one of the primary causes of discomfort in virtual and augmented reality (Lambooij et al., 2009; Hoffman et al., 2008). Several recent works (Padmanaban et al., 2017; Konrad et al., 2016; Koulrieris et al., 2017) evaluated user responses to stimuli in gaze-contingent dynamic focus VR displays and found that adaptive and dynamic focus in near-eye displays can mitigate VAC. Similar dynamic focus capabilities are shown to increase task performance in an AR setting in Section 6.1.

### 2.2.2 Panum's Fusion Area

For a given object depth, there exists a range of eye vergence where binocular fusion will occur. This range is known as *Panum's fusion area* (PFA) and it is depicted in Figure 2.8 top. It is not in the scope of this work to fully describe the topic; the interested reader is directed to the work of Schor et al. (1984) and Ogle (1932). Any vergence error larger than  $0^{\circ}15'$  to  $0^{\circ}30'$  will cause a failure in binocular fusion resulting in either suppression or diplopia, while smaller errors will lead to a loss in stereoacuity (Hoffman et al., 2008).

When presenting virtual stimuli, any inconsistent error in horizontal position of the separate eye's virtual images will result in incorrect depth perception rather than failure in binocular fusion as the user adapts their vergence to the displayed stimulus. However if unmatched intra-eye distortion occurs beyond the range of PFA, binocular single vision is lost.

### 2.2.3 Depth of Field / Depth of Focus

The perceptual sensitivity to lack of focus is characterized by the eye *depth of field* or *depth of focus* (DOF), which denotes the maximum range of retinal defocus that does not cause perceivable blur. For a given focal depth, there exists a range of eye accommodations where defocus blur is imperceptible. Many studies have been performed in an attempt to characterize the size of the DOF for the human eye and the factors which affect it.

In the most crude of simplifications, many display makers simplify this to a single value of  $\pm 0.3$  D, based on the work of Campbell (1957), which works well as a median value. However, the DOF depends on many factors such as the pupil size, the quality of eye optics, as well as properties of observed content in terms of color and contrast, luminance levels, and spatial frequency content (Wang and Ciuffreda, 2006). The values around 0.3 D which have been reported typically only apply to DOF at the fovea for suprathreshold contrasts and photopic lighting conditions. Note that the accommodative system responds to much smaller focal changes of 0.12 D, but those are not perceivable (Kotulak and Schor, 1986). With increasing retinal eccentricity the DOF increases significantly (Ronchi and Molesini, 1975; Wang and Ciuffreda, 2004). Even in the



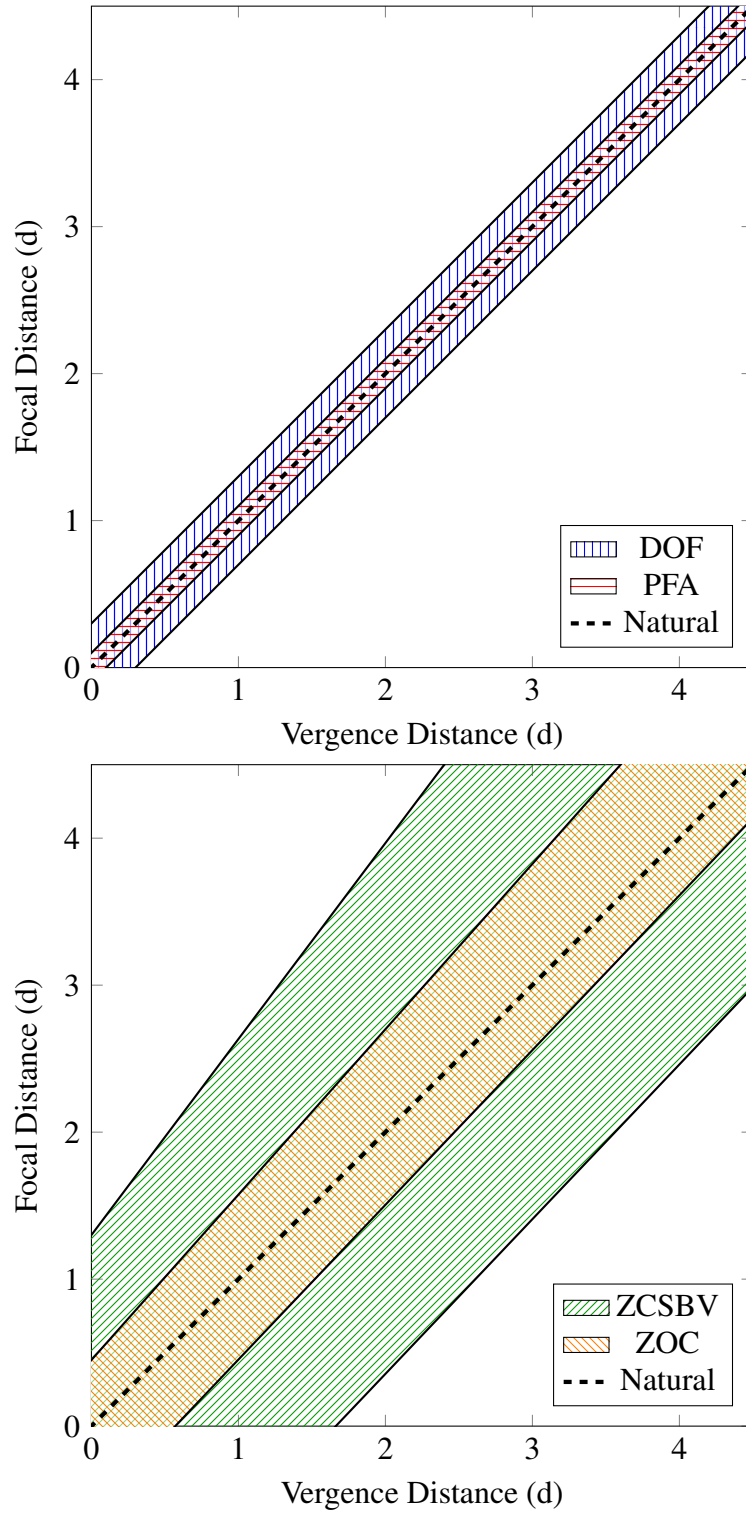


Figure 2.8: Visualizations of Panum's Fusion Area (PFA) compared with Depth of Field (DOF) [top] and the Zone of Clear Single Binocular Vision (ZCSBV) compared with the Zone of Comfort (ZOC) [bottom]. Adapted from Shibata et al. (2011).

relatively near retinal periphery of  $5^\circ$ , it increases to 2.5 D. DOF saturates at the level of 6 D to 7 D for eccentricities larger than  $30^\circ$ .

#### **2.2.4 Zone of Clear Single Binocular Vision**

For any combination of accommodation and vergence depths, there is a zone where binocular fusion and proper focus can occur. This is known as the *zone of clear single binocular vision* (ZCSBV) and is depicted in 2.8 bottom (Hoffman et al., 2008; Shibata et al., 2011; Fry, 1939). While the size of the ZCSBV varies on an individual basis, generally the boundaries are  $\pm 1.5$  D to 2 D from the natural viewing case.

#### **2.2.5 Zone of Comfort**

While it is possible to fuse images inside the ZCSBV, certain combinations of accommodation and vergence put undue stress on the visual system. Any time spent in these regions will accumulate eye strain leading to visual discomfort and fatigue. However, a subset of the ZCSBV exists where it is safe to remain for extended periods without accumulating strain; this region is called the *zone of comfort* (ZOC) which is about one-third the size of the ZCSBV (Shibata et al., 2011; Hoffman et al., 2008; Percival, 1910; Sheard, 1934).

#### **2.2.6 Accuracy of Vergence and Accommodation**

Under normal conditions, the vergence and accommodation response occur in unison. If a new stimulus is presented at a depth different from the user's current fixation point, there are three stages of reaction: (1) latency before reaction, (2) initial adjustment, and (3) correction and fine tuning. Interestingly, while the vergence angle adjusts completely to the new depth, the accommodation will lag, settling between the initial focal depth and the stimuli's focal depth (Wang and Ciuffreda (2006); Maxwell et al. (2010)). This is likely due to the strict bounds of PFA driving the vergence, while the looser DOF bounds drive the accommodation (Fincham and Walton, 1957; Sweeney et al., 2014).

### 2.2.7 Speed of Vergence and Accommodation

An initial reaction time, or latency, in the range of 150 ms to 200 ms for vergence and 300 ms to 500 ms for accommodation has typically been observed before the actual change is initiated (Campbell and Westheimer, 1960; Morgan, 1968; Phillips et al., 1972; Schor, 1992; Heron et al., 2001; Bharadwaj and Schor, 2005). While Phillips et al. (1972) have observed latencies as short as 200 ms for accommodation, the probability of their occurrence is very low. They hypothesize that such short latencies can be explained by coincidence or confusion of some subjects who have not carefully followed the experiment protocol.

Response time is dependent on the distance traveled and the direction of the fixation change, either far to near, meaning accommodation and convergence, or near to far, meaning disaccommodation and divergence. The duration of convergence and divergence are typically within 200 ms to 800 ms and lens accommodation is in the range of 500 ms to 800 ms (Campbell and Westheimer, 1960; Phillips et al., 1972; Bharadwaj and Schor, 2005; Heron et al., 2001; Morgan, 1968; Schor, 1992; Erkelens et al., 1989).

The velocity of accommodation is a useful measure of the lens accommodation dynamics. Bharadwaj and Schor (2005) observed a smooth increase in velocity to its peak value and then its slightly slower reduction to a steady state. As accommodation magnitude increased, so did the peak velocity – with a maximum value of around 10 D/s. Kasthurirangan et al. (2003) observed a similar average peak velocity for the lens accommodation, but a high variance can be observed in their data. Also, for disaccommodation, peak velocities over 20 D/s have been measured for the large accommodation magnitudes of 4 D to 5 D.

Erkelens et al. (1989) report a maximum vergence velocity of convergence of  $120^{\circ}/s$  to  $190^{\circ}/s$  and divergence of  $160^{\circ}/s$  to  $180^{\circ}/s$  for  $32^{\circ}$  vergence amplitudes. Mean vergence velocities for the same amplitudes were reported as  $38^{\circ}/s$  to  $59^{\circ}/s$  for convergence and  $45^{\circ}/s$  to  $53^{\circ}/s$  for divergence.

Interestingly, when performed in conjunction with other eye motions, such as saccades, both the latency reaction and the response duration of accommodation are decreased, while the re-

sponse duration of vergence is also decreased. Schor et al. (1999) report that the period of latency for accommodation is reduced by 13 % and the velocity of accommodation response increased by 27 % when a simultaneous saccade is performed. Enright (1984) reports that vergence is sped up by between 30 % and 40 % when synchronous saccadic motion accompanies the depth change.

All expressed values for the dynamics of vergence and accommodation are for the case when a subject is presented with an unanticipated stimulus. However, the case when voluntary depth change occurs, as with the user-driven fixation in varifocal optical see-through (OST) AR displays, has not fully been studied. Some discussion on the topic by Ciuffreda and Kruger (1988) indicates that under the user-driven condition an estimated amount of accommodation, which could be used to resolve details, would occur before any retinal-blur information is received. This preprogramming would make the visual scanning process more efficient. Kruger and Pola (1986) suggest that the natural mode of the accommodation system is anticipation, while Kruger and Pola (1987) show that accommodative prediction is a key component in tracking regularly approaching and receding objects. These discussions indicate that for voluntary depth changes, the latent period would be drastically reduced, if not eliminated entirely, and pre-accommodation may occur, such that the response time would happen partially before the fixation change is initiated. Unfortunately, until further studies are made on user-driven changes to fixation depth, we must base our requirements on reported values, and therefore use the stimulus-driven results described above.

## **2.3 Considerations in Developing Near-Eye Displays**

There are many challenges and considerations in developing NEDs. Wide-spread usage of AR will likely require a device that can seamlessly integrate into daily activity. It is speculated that such a device must be non-encumbering, yet allow natural input using our body and voice, and full-fidelity output using a combination of our senses. Specifically for vision, it would require a display equally or less intrusive than prescription eyeglasses which provides superior quality virtual imagery — replicating most or all human vision capabilities.

Several limitations in reaching these goals are related to the underlying technologies such as the interplay between battery size and weight, heat dissipation, run time, and computational power, while other considerations arise from trade-offs between features and capabilities of the device design. If not considered carefully, the design may prevent users from having comfortable, long-lasting experiences. In this section, I will focus on device design considerations with bearing on the generation of the visual imagery while attempting to discuss the applicable principles and choices required when designing a head-mounted display (HMD).

There are a few attributes which should be minimized as much as possible. Those include:

- Size – the physical dimensions of the device. Since these displays are head-mounted, smaller is better.
- Weight – the physical weight of the device. The mass mounted on the head needs to not only be minimized, but consideration should be given to its distribution.
- Computational Power – the amount of computation dedicated to forming the optical image. Primary rendering of the scene is not included. A simple example is the computation to generate the binary frames from a DMD light engine which are time-integrated to form higher contrast and full-color images.
- Operational Power – the amount of power required to generate and transport the photons to the eye. For mobile devices which operate from battery power, maintaining a low power draw is essential for extended usage.
- Heat Generation – the amount of heat generated by the optical system. For devices mounted to the head, minimizing heat generation is very important.

There are many attributes which should be increased as close to the human perceptual boundaries described in Sections 2.1 and 2.2 as possible. Those include:

- Field of View (FOV) – the solid angle for the largest displayed image as seen by the user. For simplicity, it is generally reported as a single value representing the greatest diagonal

extent, but separate width and height values are sometimes given. As shown by Arthur (2000), a limited FOV degrades task performance among other effects.

- Angular Resolution – the number of cycles per degree the display is capable of displaying; measured by modulation transfer function (MTF), a measurement of contrast per spatial frequency. Consideration should be given to the resolution required by the applications for the display; in particular the presentation of text, as the correct method of font hinting in HMDs is an open problem.
- Eye Box – the area within which a view-able image is generated by the display; measured in millimeters of diameter. The light from the display must be able to enter the pupil of any user as they change gaze direction, either through adjustment or size. The range of both interpupillary distance (IPD) and eye motion should be considered.
- Focus Range – the range of depths where images may appear in sharp focus – can be time-varying; measured as the difference of diopters of optical power from near to far. To mitigate VAC and ensure images stay within the ZCSBV, it is necessary to provide accurate depth cues for the rendered synthetic imagery, which either requires recreating the entire light field accurately, or actively changing the focus of the virtual image to optically register it to the appropriate real-world depth.
- Frame Rate – the frequency at which consecutive, complete images may be displayed; measured in Hertz. Here, not only does the CFF need to be considered, but also motion-to-photon latency and frame persistence which may cause dynamic registration errors (Lincoln et al., 2016). Due to the motion effects of being head-mounted, the illumination interval timing and length must be correct. Illuminating the frame for too long, or *high persistence*, will lead to motion blur on the retina during head motion, while extremely short illumination periods, or *low persistence*, will cause a decreased brightness. Additionally, illuminating the frame too late, with a large latency, will cause image swimming which leads to simulator sickness.

- Real-world Transmittance – the percentage of real-world light transmitted through the display; measured as a percentage of transmittance. For VST systems latency should also be considered.
- Occlusion Resolution – the number of distinct volumetric real-world regions which can be dynamically blocked from transmitting light. For displays without occlusion, this is zero.
- Luminance – the average luminous intensity per unit area of light emitted from the display; measured in candela per square meter.
- Contrast Ratio – the ratio of luminance for the brightest white to that of the darkest black simultaneously displayed; measured as a ratio.
- Dynamic Range – the range of luminance for the brightest white to that of the darkest black the display is capable of displaying given time for adaptation; measured as a ratio.
- Color Gamut – the set of colors the display can accurately reproduce; measured as a percentage of coverage of CIE 1931.

While many designs will have trade-offs specific to the optical or computational nature of the display, such as the computational power/frame rate/focal range fidelity trade-off of light field displays, here I discuss trade-offs which are mostly global to all NEDs.

### 2.3.1 Étendue or Throughput

*Étendue*, or throughput, is a measure of the flux gathering capability of an optical system. When étendue  $\varepsilon$  is combined multiplicatively with the radiance  $L$  (intensity of the light) it gives the total power  $\Phi$  of the light passing through the system as defined here for a perfectly uniform and Lambertian source:

$$\Phi = L\varepsilon \tag{2.2}$$

For a lossless optical system, étendue is invariant; the light entering the system is equivalent to the light leaving the system. In a display, étendue is the product of the area of the imaging source  $A_i$  and the solid angle of light that enters the optical system  $\Omega_i$ , which is equivalent to the product of the size of the eye box  $A_e$  and the FOV  $\Omega_e$  as described by the following equation:

$$\varepsilon = A_i\Omega_i = A_e\Omega_e \quad (2.3)$$

Equation 2.3 shows the direct trade-off between FOV and eye box size. If we seek to increase both FOV and eye box, a larger imaging source is required meaning the size of the device will likely increase. So there is an interplay between the display size, FOV, and eye box. Additionally, for a given imaging source, if we increase the FOV by decreasing the eye box, the resolution of the output image will be spread across a larger angle meaning a decrease in the angular resolution.

One alternative often used to get around the limitations of étendue is pupil expansion. For pupil expansion, the radiance of the image source is split into multiple optical paths with separate exit pupils or eye boxes. In this manner, the luminance of the image is reduced, but with multiple eye boxes produced, the effective eye box size is increased.

### 2.3.2 Focal Range Trade-off

There are many ways to increase the focal range of the display, many of which are discussed in Section 2.4. They typically can be described in terms of two categories: optical and computational. The optical methods generate multiple optical light paths with differing focal depths. These optical methods may be either spatially or temporally multiplexed to provide the full focal range of the display. Spatial multiplexing increases the size of the display while temporal multiplexing will increase the computation required for synchronizing and displaying the correct images at the correct time. Computational methods of providing a large focal range may actually



be used to decrease the overall size of the display, but come at a large cost in the computation required for displaying the imagery.

### **2.3.3 Real-world Transmission**

In AR, the manner and quality that the real world is conveyed to the eye is important to consider. OST has the advantage of letting the light from the real world, in its full complexity, pass directly to the user; however some amount of attenuation and/or blockage is likely to occur. If the display is meant for outdoor use, attenuation of the bright sunlight may be desired so that the lower luminance of the display may be perceived; however indoors, the same attenuation would darken the view undesirably. Ideally the attenuation will be at a level where virtual content may be overlaid on either bright or dark backgrounds with enough contrast to provide clear viewing. The contrast of virtual content could be enhanced with an occlusion capable display, but such displays exhibit additional problems with bulk, correct occlusion focus, and occlusion latency.

### **2.3.4 Visual Quality**

The luminance, contrast, and dynamic range of the display are co-dependent; with a greater luminance, better contrast and dynamic range usually follow, so providing a greater maximum luminance is desirable. Additionally, reducing noise, as measured in peak signal-to-noise ratio, in the final image and maintaining a high MTF are important to the final image quality.

## **2.4 Prior Work**

While the importance of the disparity between two eyes has been known since ancient times, it wasn't until the parallel inventions of the stereoscope by Sir Charles Wheatstone in 1838 and James Elliot, Esq. in 1839 when stereoscopic imagery could be presented to a viewer (Wheatstone, 1838; Elliot, 1852). With the addition of lenses by Sir David Brewster, the first NED capable of presenting static images was created (Brewster, 1849, 1856). While many versions of the

stereoscope capable of recreating motion using rotating disks or wheels were developed, it was the development of projected motion pictures that enabled Emile Reynaud in 1889 to create the first stereoscope capable of displaying motion sequences of arbitrary length (Zone, 2007).

The development of electronic image capture, transmission, and reconstruction — collectively called television — enabled live-streaming video via cameras. In 1961, Philco Corporation demonstrated Headsight, the first system to employ both a real-time video stream and head motion tracking which was linked to a motion-controlled closed-circuit camera for natural views of a remote location (Kiyokawa, 2007a). The next major step for NEDs came after the development of modern computers, which enabled interactive presentation of 3D data as shown by Sutherland (1968). Despite the display not providing either stereoscopic or focal cues, it showed the potential of presenting motion-tracked computer-generated imagery on an HMD. Around the same time period, Thomas Furness III and his Visual Display Systems Branch in the United States Air Force were developing their own helmet-mounted displays for flight control and pilot training including binocular, motion-tracked, real-time systems (Furness, 1986).

From these foundations, HMDs continued to evolve from the research lab to industrial tools and eventually from experiential arcades to consumer products. Researchers have proposed several classical optical designs (Kiyokawa, 2007b; Arthur, 2000; Nagahara et al., 2006; Sisodia et al., 2005) to address improving FOV. As demonstrated by Ilie et al. (2004) and Benko et al. (2015), combining an NED with projections promises a larger FOV, but it introduces new practical challenges. Throughout this refinement process, while attributes such as form-factor, weight, and FOV were improved, it wasn't until relatively recently when increasing the focal depth range was widely addressed as a means for solving VAC.

For the rest of this section, I review designs that have enabled accommodative cues, investigate their characteristics, and provide a comparison of these solutions in Table 2.1 and Figure 2.9. The review articles from Hua (2017) and Kramida (2016) provide additional descriptions of the various focus-supporting display architectures.

Supporting accommodative cues is known to cause major complications in an NED’s optical design. Several different techniques have been shown effective in providing more than a single focal depth. Always-in-focus and extended-focus methodologies such as Maxwellian View displays discussed in section 2.4.1 are advantageous because user state does not need to be accounted for. They can imitate defocus blur during image generation, and may provide large FOV with a small form-factor, but are typically limited in angular resolution due to diffraction of the pupil. Virtual retinal displays detailed in section 2.4.2, can provide correct focal cues through beam shaping, but face a difficult trade-off with resolution and time. Multifocal approaches presented in section 2.4.3 enable correct focal cues at the depths of the virtual images, but suffer from loss of resolution when interpolating to distances between the focal planes and a limited FOV. Varifocal techniques examined in section 2.4.4 provide high angular resolution and accommodative cues, but historically suffer from limited FOV. Recent works have improved the FOV, making varifocal a good optical choice for reducing VAC while being computationally and optically simpler than other techniques. Computational methodologies such as light fields characterized in section 2.4.5 can provide accommodative cues while enabling wide FOV. However, light field displays demand a significant amount of image formation computation and are limited in angular resolution. Displays based on computer generated holography summarized in section 2.4.6 are capable of providing per-pixel focal depth, but due to current spatial light modulator (SLM) technology, have a very constrained étendue.

### **2.4.1 Maxwellian View Displays**

The procedure of projecting an illuminating source on the eye’s pupil instead of viewing it directly is called Maxwellian viewing (Westheimer, 1966). Displays that are designed to follow this principle, where the screen plane is the optical conjugate of the retina and the illumination source is the conjugate of the eye’s pupil plane, are called Maxwellian view displays. Since the source is imaged on the eye’s pupil, and the image is formed on the retina, the result is typically

---

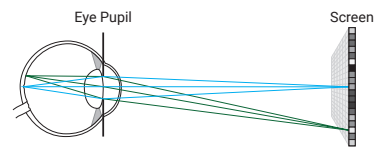
<sup>1</sup> Table 2.1: values not reported; extrapolated assuming perfect MTF for optical components and ignoring diffraction

Table 2.1: Comparison of Near-Eye Displays with Accommodative Cues Part 1

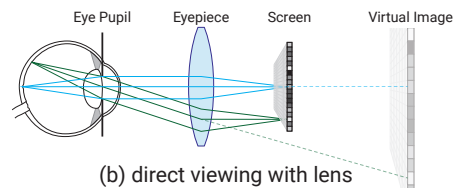
Display	Form-Factor	See-Through	FOV	Angular Resolution	Frame Rate
Medium Range Values	HMD	Limited	35-65°	8-25 cpd	45-75 Hz
Maxwellian View and Afocal Displays					
Ando and Shimizu (2001)	Bench-top	Very High	25°	<20 cpd <sup>1</sup>	60 Hz
von Waldkirch et al. (2003a)	Bench-top	No	No data	4-7 cpd	60 Hz
von Waldkirch (2004)	Glasses	Very High	31.5°	12.5 cpd	60 Hz
Yuuki et al. (2012)	Glasses	No	large	3-15 cpd	No data
Maimone et al. (2014)	Glasses	Limited	110°	2-3 cpd	60 Hz
Virtual Retinal Displays					
McQuaide et al. (2003)	Bench-top	High	70°	time-constrained	frameless
Schowengerdt et al. (2003)	Bench-top	High	No data	laser-pulse constrained	60 Hz
Jang et al. (2017)	Glasses	Very High	68°	7-9 cpd	10 Hz
Multiplane Displays					
Akeley et al. (2004)	Bench-top	No	26°	22-38 cpd	12 Hz
Love et al. (2009)	Bench-top	No	No data	27-47 cpd	45 Hz
MacKenzie et al. (2010)	Bench-top	No	18.3°	21 cpd	75 Hz
Hu and Hua (2014)	HMD	High	40°	9-12 cpd	60 Hz
Llull et al. (2015)	Bench-top	High	31°	23 cpd	67 Hz
Matsuda et al. (2017)	HMD	No	18°	3-12 cpd	60 Hz
Liu et al. (2018)	Bench-top	High	29.6°	<30 cpd <sup>1</sup>	60 Hz
Rathinavel et al. (2018)	Bench-top	High	34-52°	<10 cpd <sup>1</sup>	60 Hz
Varifocal Displays					
Liu et al. (2008)	Bench-top	High	28°	10-14 cpd	85 Hz
Konrad et al. (2016)	HMD	No	36°	5-6 cpd	75 Hz
Akşit et al. (2017)	HMD	High	60°	18 cpd	60 Hz
First Prototype	Bench-top	High	75°	No data	60 Hz
Second Prototype	HMD	High	75°	4-5 cpd	60 Hz
Light Field Displays					
Lanman and Luebke (2013)	Glasses	No	33.3°	2-3 cpd	15-70 Hz
Maimone and Fuchs (2013)	Glasses	Limited	65°	No data	85 Hz
Hua and Javidi (2014)	HMD	High	40°	10-20 cpd	60 Hz
Huang et al. (2015)	HMD	No	110°	3-4 cpd	60 Hz
Holographic Displays					
Moon et al. (2014)	HMD	High	No data	No data	No data
Maimone et al. (2017)	Bench-top	No	46°	<24 cpd <sup>1</sup>	20 Hz
Maimone et al. (2017) AR	Bench-top	Very High	>94°	<12 cpd <sup>1</sup>	20 Hz
Shi et al. (2017)	Bench-top	No	15°	<130 cpd <sup>1</sup>	20 Hz

Table 2.2: Comparison of Near-Eye Displays with Accommodative Cues Part 2

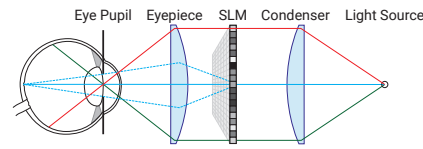
Display	Eye Box	Focal Range	Depth Switch	Blur	Eye Tracking
Medium Range Values	5-14 mm	1.5-6.0 D	15-300 ms	Rendered	Pupil-driving
Maxwellian View and Afocal Displays					
Ando and Shimizu (2001)	8×2 mm	Afocal	None	Rendered	Required
von Waldkirch et al. (2003a)	2.1 mm	7 D	None	Rendered	Pupil-driving
von Waldkirch (2004)	7 mm	>5 D	None	Rendered	None
Yuuki et al. (2012)	0.8 mm	Afocal	None	Rendered	None
Maimone et al. (2014)	7 mm	Afocal	None	Rendered	Pupil-driving
Virtual Retinal Displays					
McQuaide et al. (2003)	3.5 mm	3 D	1 kHz	Near-correct	None
Schowengerdt et al. (2003)	No data	14 D	1 kHz	Near-correct	None
Jang et al. (2017)	dynamic	3 D	None	Rendered	Required
Multiplane Displays					
Akeley et al. (2004)	≤1 mm	1.33 D	None	Near-correct	None
Love et al. (2009)	≤1 mm	1.8 D	<1 ms	Near-correct	None
MacKenzie et al. (2010)	≤1 mm	1.33 D	None	Near-correct	None
Hu and Hua (2014)	6.3 mm	3 D	1 kHz	Near-correct	None
Llull et al. (2015)	4 mm	3 D	3 ms	Near-correct	None
Matsuda et al. (2017)	No data	3.25 D	None	Near-correct	Required
Liu et al. (2018)	23.5 mm	3.05 D	1 kHz	Near-correct	Required
Rathinavel et al. (2018)	10 mm	6.5 D	60 Hz	Near-correct	None
Varifocal Displays					
Liu et al. (2008)	3 mm	8 D	74 ms	Rendered	Required
Konrad et al. (2016)	4 mm	9.5 D	15 ms	Rendered	Required
Akşit et al. (2017)	15 mm	5 D	410 ms	Rendered	Required
First Prototype	30 mm	7 D	300 ms	Rendered	Required
Second Prototype	25 mm	10 D	200 ms	Rendered	Required
Light Field Displays					
Lanman and Luebke (2013)	7.6 mm	3.3 D	None	Near-correct	None
Maimone and Fuchs (2013)	≤1 mm	9.8 D	None	Near-correct	Pupil-driving
Hua and Javidi (2014)	6.5 mm	3 D	None	Near-correct	None
Huang et al. (2015)	8 mm	4.45 D	None	Near-correct	Pupil-driving
Holographic Displays					
Moon et al. (2014)	3.14 mm	No data	None	Correct	None
Maimone et al. (2017)	≤1 mm	6.9 D	None	Correct	Pupil-driving
Maimone et al. (2017) AR	≤1 mm	No data	None	Correct	Pupil-driving
Shi et al. (2017)	10 mm	1.75 D	None	Correct	None



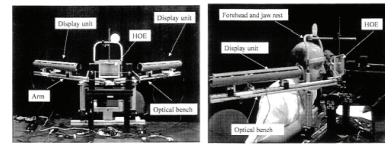
(a) direct viewing



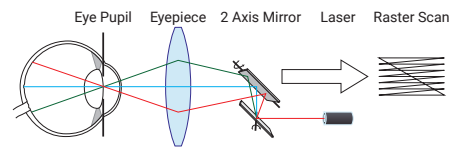
(b) direct viewing with lens



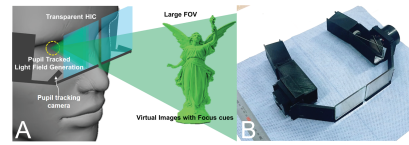
(c) Maxwellian view



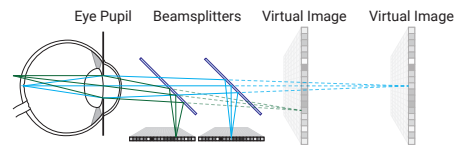
(d) Ando and Shimizu (2001)



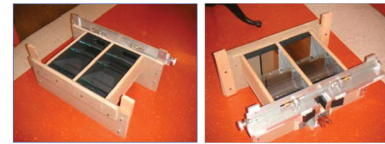
(e) virtual retinal display



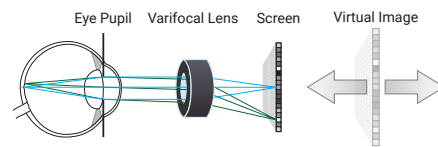
(f) Jang et al. (2017)



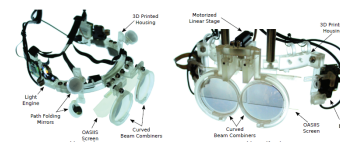
(g) multiplane



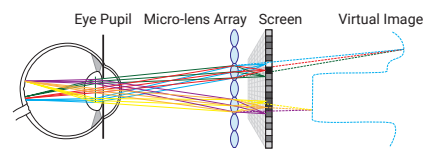
(h) Akeley et al. (2004)



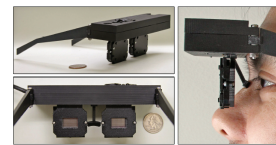
(i) varifocal



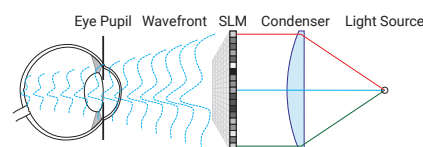
(j) Akşit et al. (2017)



(k) light field



(l) Lanman and Luebke (2013)



(m) holographic



(n) Maimone et al. (2017)

Figure 2.9: Classes of near eye displays capable of large focal range. Left diagrams show the basic optical principles for each class of display, while right photos show example displays for each class.

always in focus. Several variants of the Maxwellian view displays have been proposed in the literature. Ando et al. (1998, 2000); Ando and Shimizu (2001) utilize both a holographic optical element and a digital micromirror display (DMD) (Hornbeck, 1997) to achieve retinal projection. While von Waldkirch et al. (2003b) use an LCD (liquid crystal display) and a set of collimation and projection lenses to achieve a quasi-accommodation-free Maxwellian viewing with high resolution. However, such displays have a very small depth of focus and eye box sizes. By using an elliptical scanning beam, von Waldkirch et al. (2003a) show that the depth of focus of a Maxwellian view retinal scanning display can be improved, whereas von Waldkirch et al. (2005) show an improved depth of focus by using an oscillating fluid lens in the retinal projection system. Yuuki et al. (2012) combined an LCD, pinhole array aperture mask, and micro-lens array to create a tessellated series of Maxwellian view displays capable of extended DOF.

Alternative always-in-focus mechanisms, related to Maxwellian view in that they project an image directly on the retina, also offer sharp imagery regardless of the viewer's accommodation state. The "Pinlights" always-in-focus AR display (Maimone et al., 2014), by using a see-through sparse backlight mechanism behind an LCD, generates a tessellated series of retinal projectors capable of a wide FOV but limited in angular resolution.

#### **2.4.2 Virtual Retinal Displays**

Another method for projecting light directly onto the retina is by conditioning a narrow bundle of collimated rays and directing it into the eye in a steerable manner. By modulating the intensity and color of the bundle of rays while raster scanning the light across the retina at a high frequency, full image generation is possible. This technique is called virtual retinal display, or retinal scanning display, and was first proposed by Kollin (1993). A virtual retinal display was demonstrated in conjunction with micro-electromechanical system (MEMS) deformable mirror membrane devices (DMMDs) by McQuaide et al. (2003). Built on the principles of a Maxwellian view display, a laser is scanned onto the DMMD, which then reflects it through a series of mirrors directly into the pupil — forming an image on the retina. The surface convexity of the mir-

ror is controlled by the applied voltage, thereby controlling the focus of the displayed objects. Schowengerdt et al. (2003) showed an achievable accommodation range of 0 D to 14 D by the DMMD. Creating a volumetric display by application of deformable mirror membranes was attempted by Schowengerdt and Seibel (2006), where the membrane curvature was synchronized with per-frame swapping between two different images, thereby displaying the images at different depths simultaneously. The prototype demonstrated a depth range of 0 D to 16 D in a contiguous fashion. More recently, Jang et al. (2017) demonstrated a Maxwellian view style virtual retinal HMD with multiple projectors and a holographic optical element. While the Maxwellian view capability of the display extends the depth of field, the small eye box limitation is overcome by employing eye tracking and a moving eye box.

### **2.4.3 Multifocal Displays**

Multifocal displays are capable of generating virtual images at more than one focal depth simultaneously. In the classical multiplane approach introduced by Rolland et al. (2000), virtual content displayed at one of the focal planes will have correct focus cues, but generating virtual content between the focal planes requires interpolation leading to less-correct focus and loss of resolution. These displays have large computational demands for decomposing and interpolating the virtual content and complex optical hardware that doesn't typically lead to a wearable form-factor. Much work has been done on improving scene decomposition and gaze-contingent multiplane capabilities (Mercier et al., 2017; Narain et al., 2015). There are two approaches for generating multiplane displays: optical path multiplexing and temporal multiplexing.

Akeley et al. (2004) demonstrated the benefits of fixed-viewpoint optically-multiplexed multiplane desktop displays with a prototype capable of generating near-correct focus cues without any need for eye tracking. Love et al. (2009) used two fast switchable lenses per eye to create a time-multiplexed four-plane display. The work described in Hu and Hua (2014) demonstrates a see-through, time-multiplexed, multiplane display in the form of a wearable NED utilizing a 1 kHz DMMD. Unfortunately, such a display design offers good resolution, but only with a small



field of view. Mercier et al. (2017) developed a bench-top prototype combining a three-plane display, gaze tracking, and focus tracking to demonstrate that despite multiplane displays showing near-correct focus cues without eye tracking, correct scene decomposition is dependent upon eye position. This means that eye tracking is required for displaying high quality images. Recently Rathinavel et al. (2018) detailed an extremely fast, time-multiplexed, multiplane display supporting 280 distinct depth planes. The density of depth planes is indistinguishable from a full volumetric display and shows the great potential of  $>20$  kHz operating DMDs. However, the optical complexity in such approaches has thus far challenged their practicality in increasing FOV and decreasing form-factors.

As an alternative to multiplane displays, Matsuda et al. (2017) demonstrated a focal surface display, which uses a phase-only SLM to bend the focal plane of the image into a complex surface. These scene-optimized surfaces can improve the focal accuracy across simple scenes. They also propose combining multiple focal surface images into a multiplane focal surface display which would greatly reduce the number of focal planes required to accurately represent a scene. Akin to multiplane planes, Konrad et al. (2016) study an interesting scenario called monovision, where each eye is subjected to one focal depth, with one eye's focus being near and the other eye's focus being far. This approach leverages binocular single vision and suppression in an attempt at reducing VAC, with a loss of resolution. Detailed perceptual studies on monovision have also been conducted by Johnson et al. (2016a) and Koulrieris et al. (2017) which found that not only did viewer comfort and visual performance not improve, but monovision displays do not drive accommodation to the simulated distance meaning they do not resolve VAC.

#### **2.4.4 Varifocal Displays**

Related to multiplane displays, varifocal displays elect to show a single, but move-able focal depth. The core idea being that the human eye can only focus at a single depth at a time, so if the displayed focus can be changed fast enough and the correct depth to display is known, only this one focal depth need be displayed. A tunable lens system combined with a spherical mirror is

used in the work of Liu et al. (2008), producing a small FOV but having a good accommodation range capable of switching depths within 74 ms. The study described in Konrad et al. (2016) also takes advantage of an electrically tunable lens system as relay optics and demonstrates a similarly small FOV VR prototype. Their solution switches depth from one extreme to another within 15 ms, and provides a better accommodation range. Akşit et al. (2017) uses holographic optical elements for intermediate image formation before relaying the final image into the eye, offering a wearable form-factor with good FOV. All of the above mentioned varifocal display designs, suffer various drawbacks either in form-factor, depth-switching speed, or FOV.

Recent studies show evidence that supporting accommodative cues through a varifocal mechanism improves visual comfort (Johnson et al., 2016b) and user performance (Konrad et al., 2016). Just as with always-in-focus displays, objects not located at the current focal depth should have appropriate rendered blur to provide the appropriate focal cues.

#### **2.4.5 Light Field Displays**

Light field displays, in addition to the traditional color, intensity, and position, provide angular control of the light generation leading to capabilities for correct parallax, stereoscopic view, and multi-view. With enough angular resolution they can also depict correct focus cues in an NED. Two approaches for creating light field displays have been presented thus far, integral and multi-layer.

Integral imaging, first proposed by Lippmann (1908), places an array of micro lenses, an aperture array, or both in between the viewer and the image such that as the viewing angle changes, so does the visible image. Alternatively, a multi-directional backlight can achieve the same effect Fattal et al. (2013). Unfortunately, current implementations of integral light field NEDs sacrifice the spatial resolution for generating angular resolution. Lanman and Luebke (2013) introduced a near-eye light field display that uses an array of microlenses, resulting in a very thin and light form-factor VR NED with a good possible field of view, but with a heavily compromised resolution. Hua and Javidi (2014) demonstrate an NED for AR applications that combines recent

advancements of free-form relay optics with a computational integral imaging methodology. Unfortunately like with most designs the transition from VR to AR was accompanied by a loss in FOV. Akşit et al. (2015) uses a pinhole mask in front of an LCD to create a light field at the eye and thus increase the apparent depth of field, but at the expense of resolution.

Multi-layer light fields use stacked SLMs with a single illumination source to add the angular control of the light. This approach works to increase the resolution above the integral approach, however it suffers from a loss of contrast as light must pass multiple SLMs and has a resolution limit from the compounding diffraction. Maimone and Fuchs (2013) detailed an AR near-eye multi-layer display with no reflective, refractive, or diffractive optical elements capable of occluding the real world. While having an impressive focal range, and decent FOV, the display suffered from noise, low resolution, and poor contrast. More recently, Huang et al. (2015) demonstrated a prototype which employs two LCDs and a pair of classical magnifiers. This light field stereoscope was capable of producing a wide diagonal FOV and an improved image resolution.

#### **2.4.6 Holographic Displays**

While holographic optical elements have been used for a long time in NED designs enabling almost eyeglasses-like thin form-factor, and a very wide FOV (Ando and Shimizu, 2001; Kim et al., 2015; Akşit et al., 2017), true computer generated holographic NEDs are a very recent topic. Holography promises good angular resolution with a thin form-factor by using phase and sometimes amplitude SLMs to manipulate of the wavefront of light causing interference to generate the image. Recently, computer-generated hologram based NED designs for VR and AR have been presented by Maimone et al. (2017), showing superior quality imagery, the ability to provide per-pixel focus, wide FOV and eyeglasses form-factor. For such displays, however, a small eye box, large compute demand, and theoretically limited resolutions still remain major concerns. Shi et al. (2017) demonstrated a real time rendering pipeline for computer-generated holograms using spherical waves and achieving high resolution and a much wider eye box.

## 2.5 Discussion

In attempting to recreate the illusion of a window into a virtual world, our goal is to only provide enough visual information to meet the maximum requirements of the HVS. While the relatively high requirements of  $\sim 60$  cpd and  $\sim 0.3$  D are the resolution and DOF targets for the foveal region, as eccentricity increases, resolution and DOF can be far less strict as sensitivity decreases. In general, the light color, intensity, and contrast should be a good match for the environment and providing good quality imagery. Consideration should be made in making the device as comfortable as possible including limiting the size and weight, increasing the battery life, and providing correct stereoscopic vergence and accommodation cues to maintain viewing inside the ZOC and even within the correct DOF if possible. As evidenced by the previous work discussed here, attaining all of these objectives is a difficult task: some concessions must be made by the display creators as they seek the best possible display for their chosen applications.

## **CHAPTER 3: DEFORMABLE BEAMSPLITTERS – A NEW OPTICAL ELEMENT**

Setting out to design and produce a large-focal-range-providing AR NED, we examined the existing display categories. The simplicity of varifocal and time-multiplexed multiplane displays, it seemed, would provide a solid foundation from which to build. In order to improve on the existing work, we set specific goals for having a wide FOV and good real-world transparency. Additionally, we wanted to maintain a fast focal change speed providing either a time-multiplexed multiplane display or, at slower speeds, a varifocal display capable of changing focus faster than the human eye. With these goals in mind, we sought an optical element that would meet our requirements:

1. fast focus changing,
2. wide FOV, and
3. good see-through characteristics.

### **3.1 Previous Work in Deformable Membrane Mirrors**

An optical device capable of providing both fast focal change and, in a large aperture configuration, a wide FOV already existed: deformable membrane mirrors. While related, these are not the MEMS-based deformable mirror membrane devices (DMMDs) frequently used to condition the laser beams in virtual retinal displays or other applications. Deformable membrane mirrors are large varifocal mirrors consisting of thin, reflective membranes stretched across an aperture and affixed to an actuation housing. They are often actuated by modulating the pressure inside the air-tight housing through use of a loudspeaker.

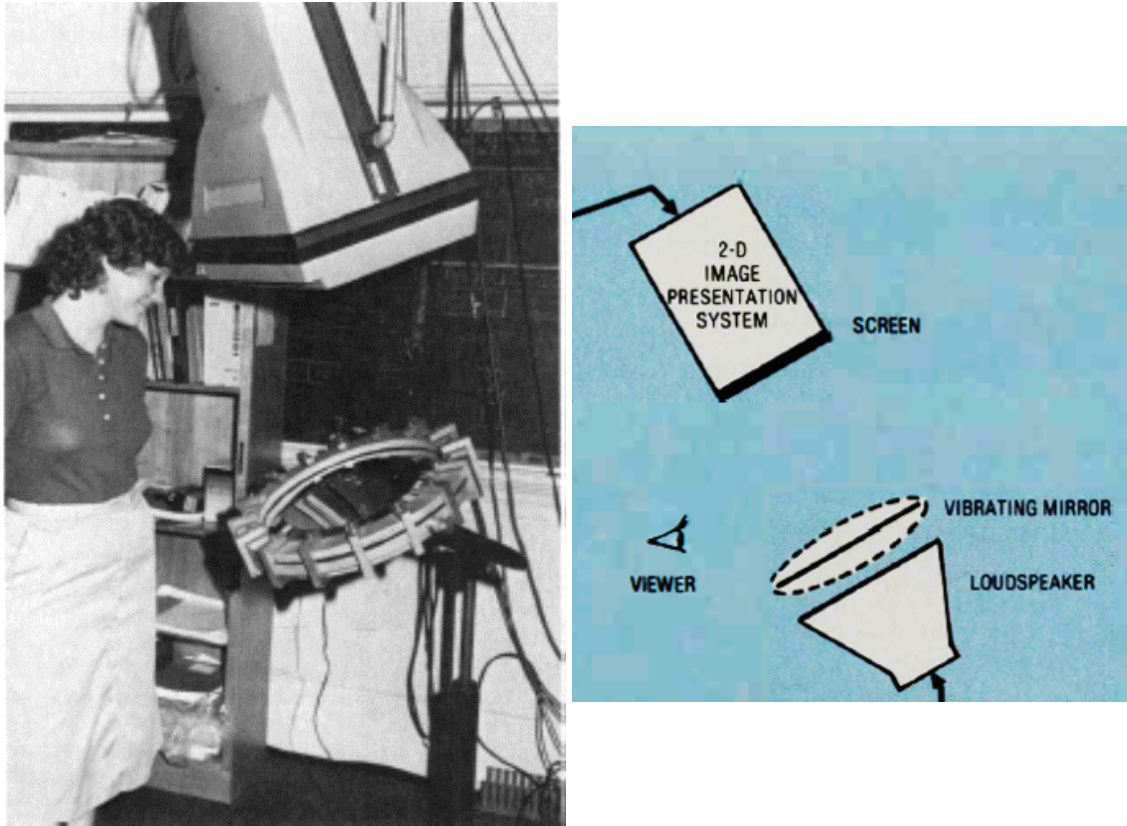


Figure 3.1: Example photograph and diagram of a deformable membrane mirror display. Used with permission from Fuchs et al. (1982)

Deformable membrane mirrors were first presented by Golay (1947) as a component in a pneumatic system for detecting infra-red radiation. Their use as variable focal length mirrors, which provided inexpensive construction and favorable size to weight ratio, was described by Muirhead (1961). Traub (1967) showed the first volumetric display system which employed deformable membrane mirrors using rapid oscillations. By reflecting either a time-synchronized oscilloscope or a cathode ray tube, the 2D display image was swept through space, generating a 3D image with correct parallax and focal cues. Additional display systems were presented by Rawson (1968), Rawson (1969), and Hobgood (1970). A system for capturing, transmitting, and displaying a 3D image by using bird-bath optics with a varifocal membrane mirror for capture, and a similar optical configuration with  $180^\circ$  out of phase membrane mirror for display was described by King and Berry (1970). Further work was performed on improving the image generation, rendering techniques, and real-time interaction by Cohen (1979), Fuchs et al. (1980),

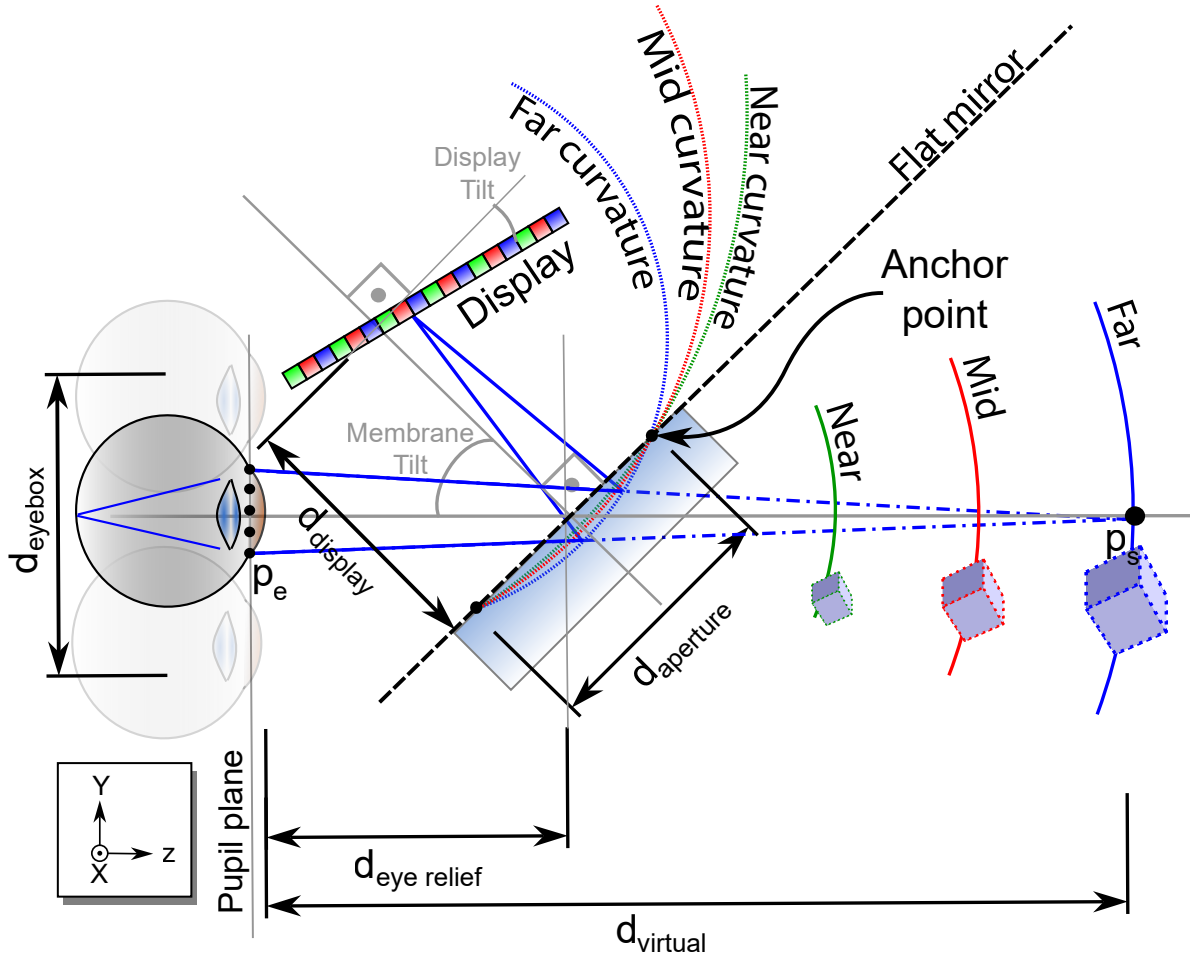


Figure 3.2: A sketch showing our deformable beamsplitter display optical layout with parameters for the single eye case. An image on a display above the user’s eye reflects from our deformable membrane beamsplitter toward the eye. A virtual image can be created at a desired position in space by varying the curvature of our deformable membrane beamsplitter. Anomalous perspective is demonstrated via the virtual cube size: display magnification increases as the distance increases.

Fuchs et al. (1982), and Mills et al. (1984). Later, McKay et al. (1999a), McKay et al. (1999b), and McKay et al. (2000) created a large 1.2 m membrane mirror capable of displaying 2D and 3D objects and enabling a telepresence system.

### 3.2 Developing Our Early Prototypes

As stated above, deformable membrane mirrors provided the first two of three requirements for improving multiplane or varifocal NEDs. By simply modifying them to be transparent, we

would have the optical element we were looking for. In achieving transparency, two obstacles presented themselves: the opacity of the housing and driving mechanism, and the opacity of the mirror. By making the rear surface of the air-tight housing transparent using acrylic and moving the actuation mechanism away from the main chamber, we quickly overcame the first obstacle. Modifying the mirror was not so simple: we sought to replace the deformable mirror with a beamsplitter capable of deforming to the needed curvatures while combining the unaltered, transmitted real-world light with the focus-changing, reflected virtual images. After several material trials, we eventually found a material capable of the requisite deformations and having the reflection and transmission characteristics desired. Combining both newly transparent portions, we had a new optical element that met all of our requirements which we named the *deformable beamsplitter*.

In developing a usable prototype capable of generating high-quality images and controlling the membrane curvature at high speeds, we performed much experimentation in membrane materials and housing construction. Most membranes described in the deformable membrane mirror literature consist of a metallized Mylar®, also known as polyethylene terephthalate (PET), film tensioned across a circular aperture. While optically-clear versions exist, at the time of investigation it was unclear if we could control the reflectance to transmission ratio. Eventually, when producing our in-house solution we developed a process of controlling the ratio, however we had already moved on to other materials and did not return to PET. Further investigation on using PET as a quickly oscillating beamsplitter would be a promising line of research.

### **3.2.1 Membrane**

Our first prototypes mainly used 0.004” clear polyvinyl chloride (PVC) film for the membrane. Unfortunately the micro surface structure of the material did not allow high quality transmission or reflections, as can be seen in Figure 3.3. A prototype display using a PVC membrane can be seen in Figure 3.4. As we sought a better material, we tried an ethylene tetrafluoroethylene (ETFE) film and 3M VHB clear acrylic tape before choosing polydimethylsiloxane (PDMS).



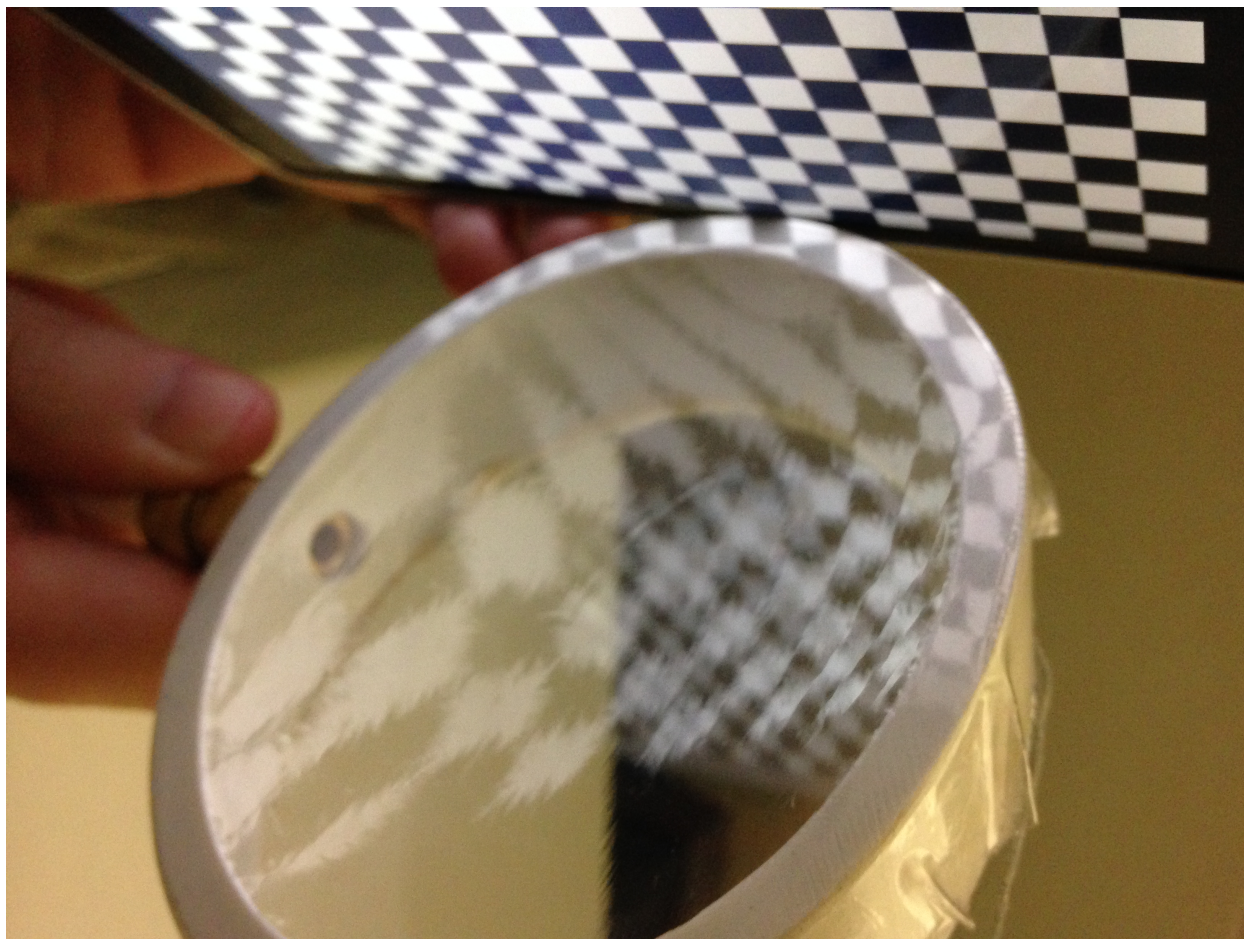


Figure 3.3: View of the poor image quality produced by the PVC film. Both image reflection (checkerboard) and image transmission (bright to dark edge) quality can be seen.

PDMS was advantageous because of its outstanding transparency in visible wavelengths, high elasticity, and providing in-lab fabrication for experimentation in formulation and membrane thickness. Metalization improves the reflection of the image as can be seen in Figure 3.5. The optical quality of both the reflection and transmission from a PDMS membrane can be seen in Figure 3.7.

The task of manufacturing custom flexible membranes is accomplished traditionally through surface micromachining, bulk micromachining, liquid crystals, piezoelectric or electrostrictive actuators as reviewed by Mansell et al. (2002). Pneumatic based systems have also been demonstrated for building tunable microoptics using PDMS (Werber and Zappe, 2008), avoiding the use of high voltages or external fields in operation and precise machining in manufacturing. On the

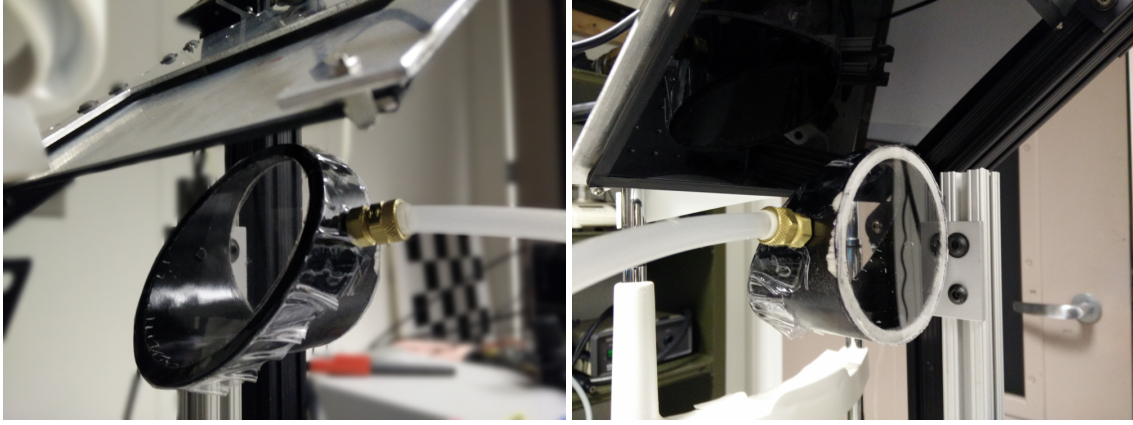


Figure 3.4: Three-quarter front and rear views of an early prototype display using a PVC tube housing with a clear PVC film membrane.

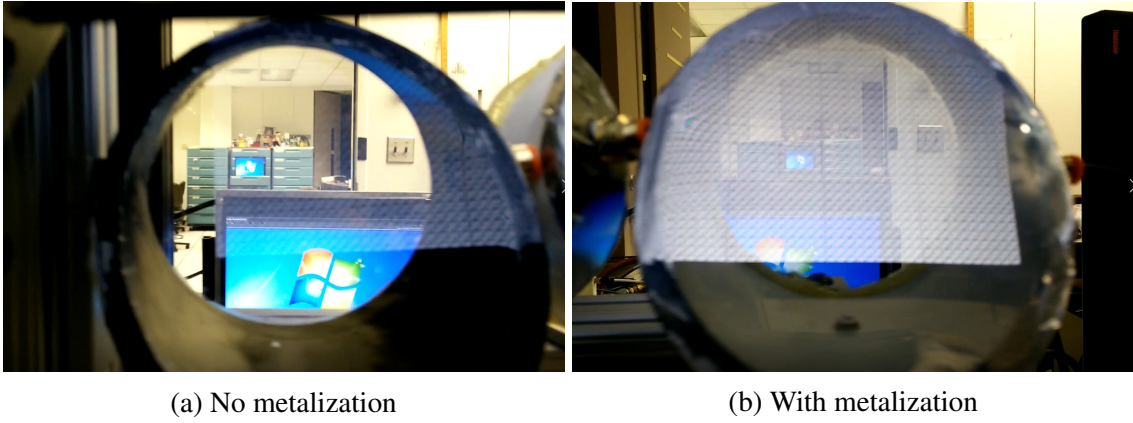


Figure 3.5: Comparing the difference in reflection intensity after membrane metalization on a PDMS membrane. In these images, the reflected image is not in focus; better quality reflections may be seen in later figures.

other hand, PDMS has numerous attractive material properties such as outstanding transparency in visible wavelengths, high elasticity, and excellent temperature stability. Our choice of a high elasticity material meant that deformations could occur without buckling, however the fast oscillations needed for a multiplane approach were unachievable, so we focused on a varifocal approach. Inspired by these advantages, we created our own recipe for the task.

In our first membrane fabrication procedure which was developed by Cary Tippets, a collaborator in the University of North Carolina at Chapel Hill (UNC-CH) Nanoscale Optical Materials lab, we used Sylgard 184 PDMS purchased from Dow Corning. Sylgard 184 is a two-part elastomer kit, with PDMS pre-polymer and a cross-linking agent. The prepolymer is mixed with

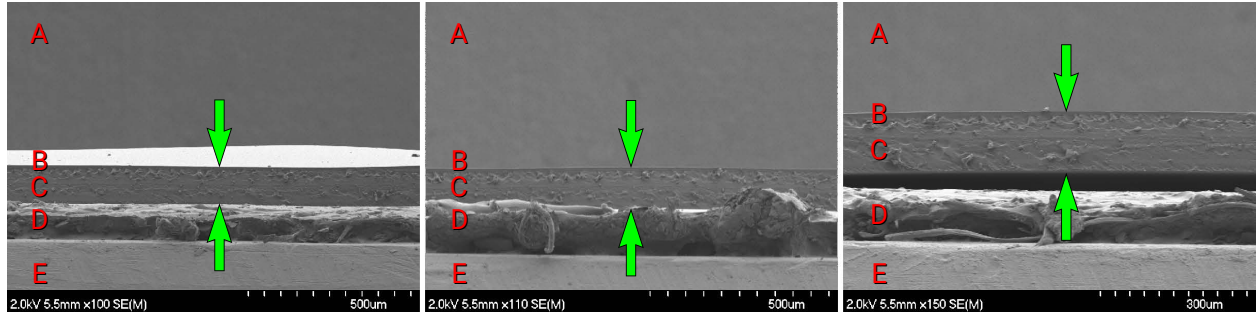


Figure 3.6: Scanning electron microscope view of membrane thickness and surface quality before sputter deposition. Sample scanned at three regions of the membrane: half of the radius (left), the center (middle), and close to the edge (right). Labels indicate: Vacuum (A), surface of the membrane (B), membrane (C), copper conductive tape (D), and metal stage (E). Smooth surface and consistent thickness of  $120\text{ }\mu\text{m}$  measured across all regions of the membrane. Arrows indicate the membrane.

cross-linking agent at a ratio of 10 : 1 and mixed vigorously for 3 min. The mixture is then degassed for 15 min, to remove bubbles incorporated during mixing. 150 mm silicon (Si) wafers were purchased from University Wafers. The wafer is silanized, to ease membrane release, by being placed in a desiccator, with  $20\text{ }\mu\text{L}$  of trichloro (1H,1H,2H,2H-perfluorooctyl) silane and evacuated for 30 min and left under vacuum for 1 h. Mixed and degassed PDMS prepolymer is spin cast on the Si wafer for 1 min at 300 revolutions per minute (rpm) to obtain a PDMS membrane of approximately  $240\text{ }\mu\text{m}$ . The membrane is placed in an oven at  $100\text{ }^{\circ}\text{C}$  for 24 h to produce a repeatable Young's modulus (Seghir and Arscott, 2015). The membrane is then placed in a commercial physical vapor deposition unit (Kurt Lesker PVD 75) and a 20 nm silver film is sputtered on the membrane. After metalization the film is carefully peeled and stretched taut across the vacuum housing to form the deformable membrane mirror as described below. This is the recipe used for the first prototype described in Section section: VacuumActuated.

After our first prototype, we modified the membrane fabrication procedure. Overall it follows the same process as described above with some alterations to improve the process and membrane quality. The silanized wafers are transferred to the cleanroom for the fabrication process, reducing particulate inclusions. Just as before, we use a Sylgard 184 PDMS kit purchased from Dow Corning. After mixing the prepolymer and cross-linking agents, the mixture is now degassed for 40 min to remove bubbles introduced during mixing. The mixed and degassed PDMS prepoly-

mer is spin cast on the silicon wafer for 1 min at a faster rate of 600 rpm reducing the thickness of the membrane to 120  $\mu\text{m}$ , thereby decreasing the pressure required to deform the membrane. The membrane is then cured and a layer of silver is vapor deposited as previously reported. After metalization, the film is carefully peeled, and using a custom designed apparatus, stretched uniformly and affixed to the vacuum housing to form the deformable beamsplitter as described below. Our membranes were fabricated with the assistance of both Cary Tippetts and Qian Dong also a member of the UNC-CH Nanoscale Optical Materials lab.

Since our display has only one optical element, it is essential to determine and maintain the quality of the membrane. With this goal, we imaged the profile of the membrane to measure the consistency across the spin-cast surface as seen in Fig. 3.6. It can be seen that the membrane from outer edge through the half-radius to the center has a consistent profile and a smooth, flat surface.

### **3.2.2 Housing**

The housing construction presented a different set of problems. We needed as close to an air-tight housing as possible while maintaining transparency and the ability to affix the membrane evenly and securely. As seen in Figures 3.3 - 3.7, we started by using a cross-section of a PVC pipe with a clear Lexan<sup>TM</sup> polycarbonate sheet affixed to the back face and the membrane on the front face. This worked well for initial trials; however it became clear that a more customizable housing shape was required, so we turned to 3D printing.

Our first 3D printing method was fused deposition modeling (FDM) which enabled fast prototyping of customized shapes as seen in Figure 3.8. However due to the construction technique, FDM printed materials are not air-tight, which hampered the pneumatic operation of our system. We attempted several methods of post-processing the FDM printed housings including acetone baths, heat treatments, and surface coatings, none of which satisfactorily provided air-tightness, so a different method of 3D printing was sought.



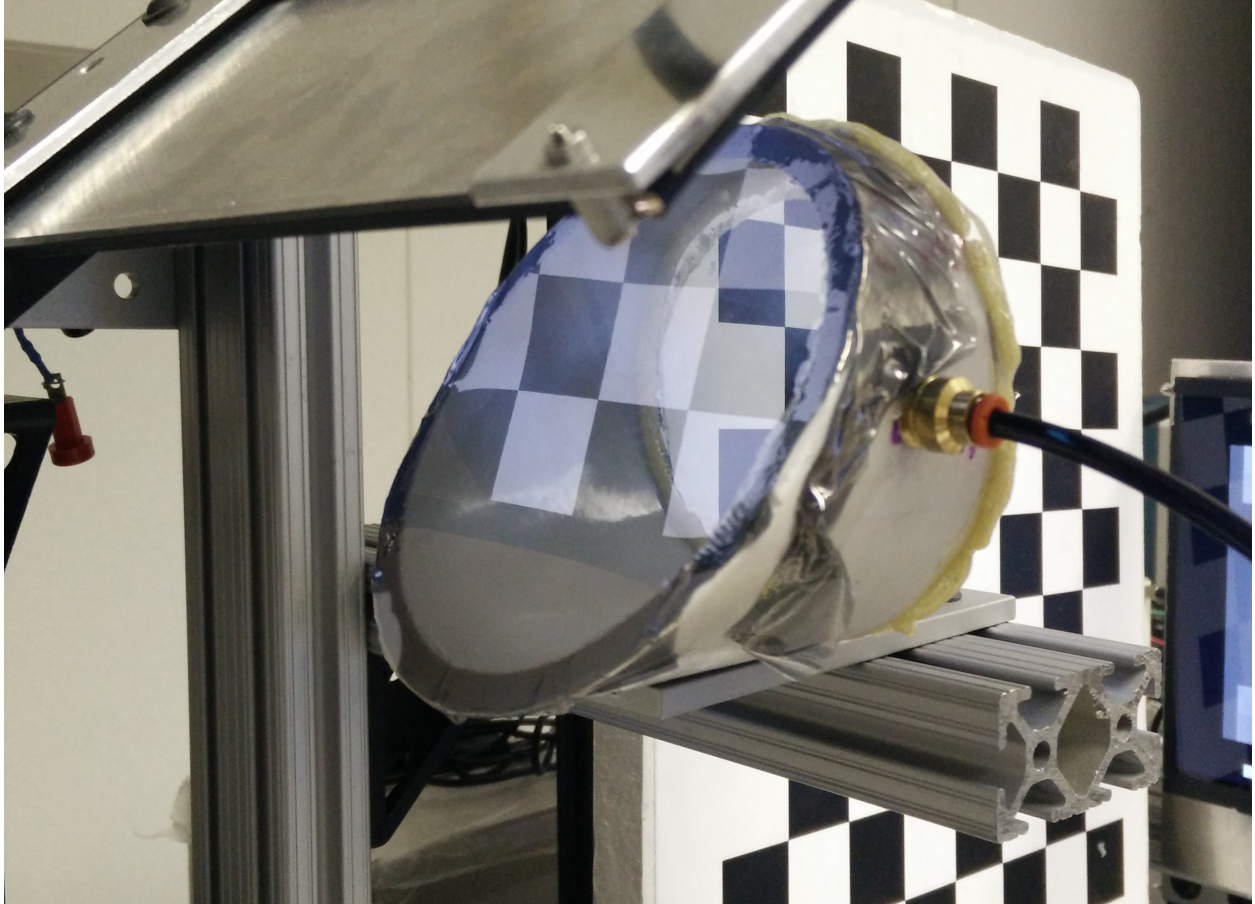


Figure 3.7: Three-quarter view of an early prototype display with PDMS membrane. The quality of both the transmitted real-world image and the reflected virtual image can be seen. Note the sharp, in-focus checkerboard patterns visible in reflection and transmission.

Our next housing was fabricated by a PolyJet 3D printer and instantly fixed all problems with being air-tight as seen in Figure 3.9. Unfortunately the cost and turn-around time made this method of fabrication prohibitive, so we turned to stereolithography apparatus (SLA) 3D printing which was able to maintain the air-tight qualities while being cheaper and faster to produce. SLA 3D printing became the final method of fabrication for both the housing and other custom pieces needed for prototype assembly. We fabricated the housing using a Formlabs 2 3D printer<sup>1</sup>, which uses SLA 3D printing with liquid photopolymer resins.

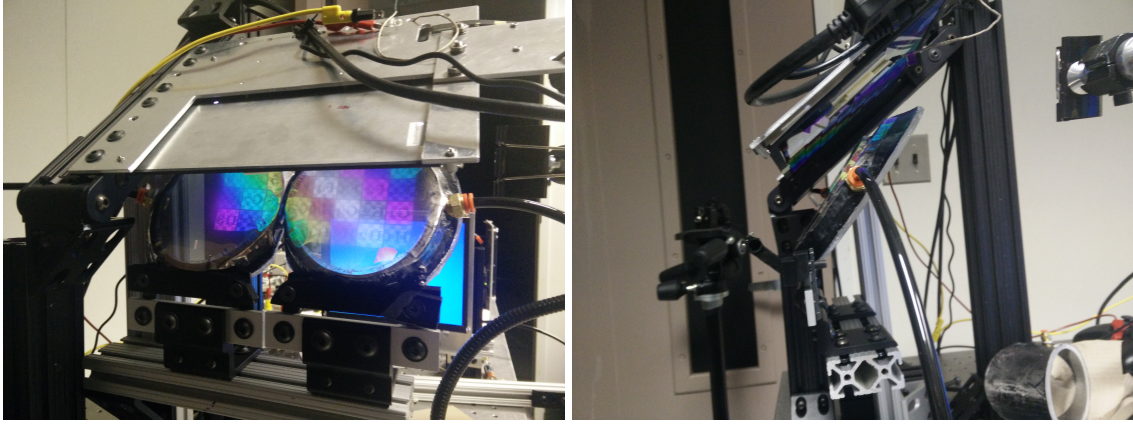


Figure 3.8: Front and side views of an early prototype display using a pair of FDM 3D printed housings.

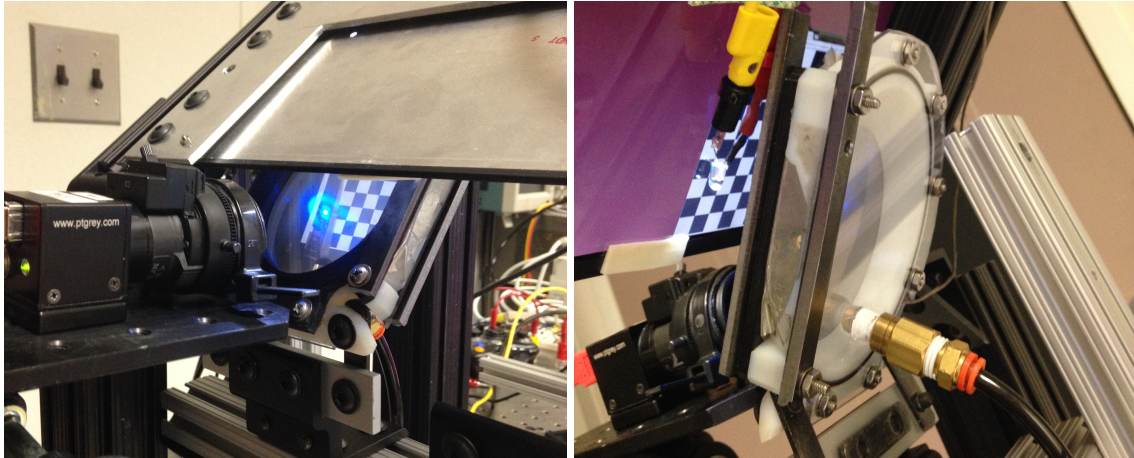


Figure 3.9: Front and side views of an early prototype display using a PolyJet 3D printed housing.

### 3.2.3 Assembly

As our experiments progressed, it became apparent the method of affixing the membrane to the housing was important, not only to create an air-tight seal and to avoid future detachments and tearing, but to ensure homogeneous isotropic strain. After transitioning to PDMS, our membrane was first attached as close to rest state as possible while ensuring a flat attachment. Problems with uneven distribution of metallization fracturing lead us to pre-stretching the membrane before attachment and re-applying a strain during attachment.

---

<sup>1</sup> <http://formlabs.com/>

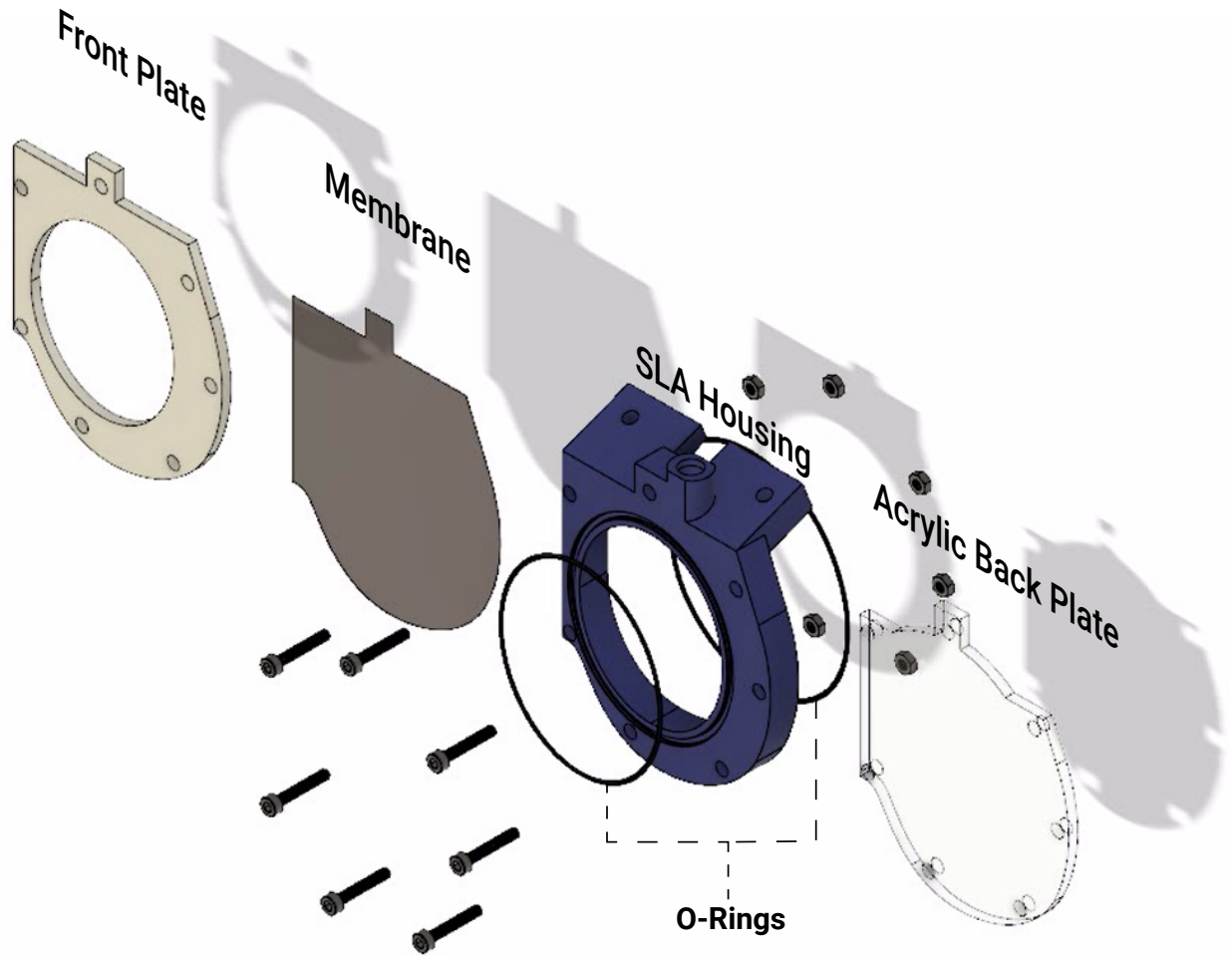
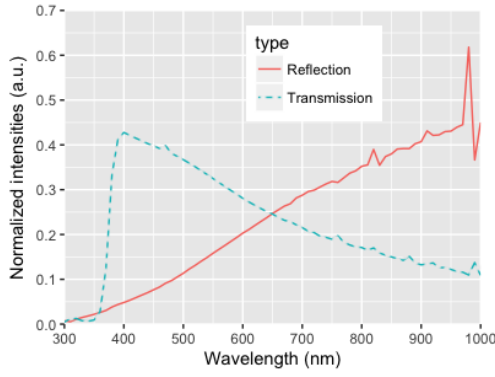


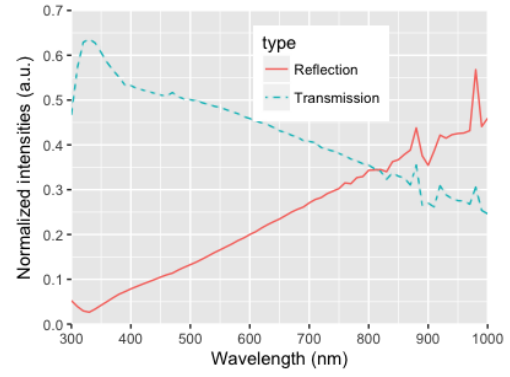
Figure 3.10: An exploded diagram illustrating the components of a deformable beamsplitter housing from prototype 2. The parts from front to back are a front plate for securing the membrane in place and ensuring a good o-ring seal, the deformable membrane beamsplitter, SLA 3D printed housing, and a see-through acrylic back plate closing up the air-tight chamber with a good o-ring seal. Screws and nuts are also depicted.

Performing the re-stretching and attachment procedure by hand lead to uneven curvature in the deformed shape of the membrane, so we devised methods and mechanisms to create homogeneous isotropic strain. Our first mechanism was an expanding pulley, however our mechanical design proved faulty, so we switched to a simpler concentric ring design. By affixing the membrane to one ring without any strain applied, then pressing the membrane over a ring with smaller radius, we were able to evenly stretch the membrane before affixing it to the housing with double-sided tape and securing it in place with an additional plate and o-ring.





(a) Reflection and transmission profile of resting-state membrane at  $21^\circ$ .



(b) Profile of stretched and housing-attached membrane at  $21^\circ$ .

Figure 3.11: Figure showing the wavelength dependent transmission and reflection characteristics of the in-house manufactured deformable membrane.  $21^\circ$  membrane tilt comes from prototype 2.

### 3.3 Properties of Deformable Beamsplitters

#### 3.3.1 Optical Properties

Transmission and Reflection characteristics of our deformable membrane beamsplitter were captured as seen in Figure 3.11 using a J. A. Woollam variable angle spectroscopic ellipsometer. The deformable membrane beamsplitter was aligned and the incident angle was set to  $21^\circ$ , matching our second prototype for both the transmission and reflection measurements. We measured the reflectance and transmissive properties of the membrane for both a non-attached membrane and a membrane stretched and attached to our housing. It is worthwhile to note the minimal stretching we perform while attaching the membrane improves the transmission characteristics while only slightly decreasing the reflection strength.

Figure 3.6 shows a scanning electron microscope view of a cross-section for a typical membrane, illustrating the thickness and surface qualities of the fabricated membranes. The work of Lee et al. (2016) highlights that a thickness of an optical combiner plays a crucial role in depth perception; as our membrane mirror has  $120\text{ }\mu\text{m}$  thickness, the effects described by Lee et al. are expected to be at a negligible level in our implementation.



As explained in Hobgood (1970), the focus of the reflection for a flat membrane beamsplitter is at infinity; however if a perfectly spherical deformation is applied, the focal point moves according to the equation

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q} = \frac{16h}{d^2} \quad (3.1)$$

where  $f$  is the focal point,  $p$  is the object distance,  $q$  is the virtual image distance,  $h$  is the distance of membrane deformation, and  $d$  is the diameter of the membrane aperture. This equation leads directly to

$$f = \frac{d^2}{16h} \quad (3.2)$$

for determining the focal point and

$$q = \frac{fp}{p-f} = \frac{d^2 p}{16hp - d^2} \quad (3.3)$$

to determine the virtual image distance for an object at  $p$ . A more thorough analysis for NED applications is presented in Section 3.4 below.

With the increased focal power needed to move virtual images further distant, more optical magnification results, meaning that as objects are focused further away, they appear larger. This phenomenon is called anomalous perspective. Additionally, as with all simple lenses, some amount of image distortion is exhibited in the spherical mirror. Both of these optical distortions must be pre-corrected by the system generating and producing the images.

### 3.3.2 Deformation Properties

In creating a stress-strain relationship model for elastic materials, Wall (1942) introduced a successful method for simulating a cross-linked 3D network of long-chain molecules based on the kinetic theory of elasticity, which was applied to the simplest forms of deformation: elongation, unidirectional compression, and shear. Later, Treloar (1944) extended the model to the general type of homogeneous deformation for elastic materials. While many models have since been

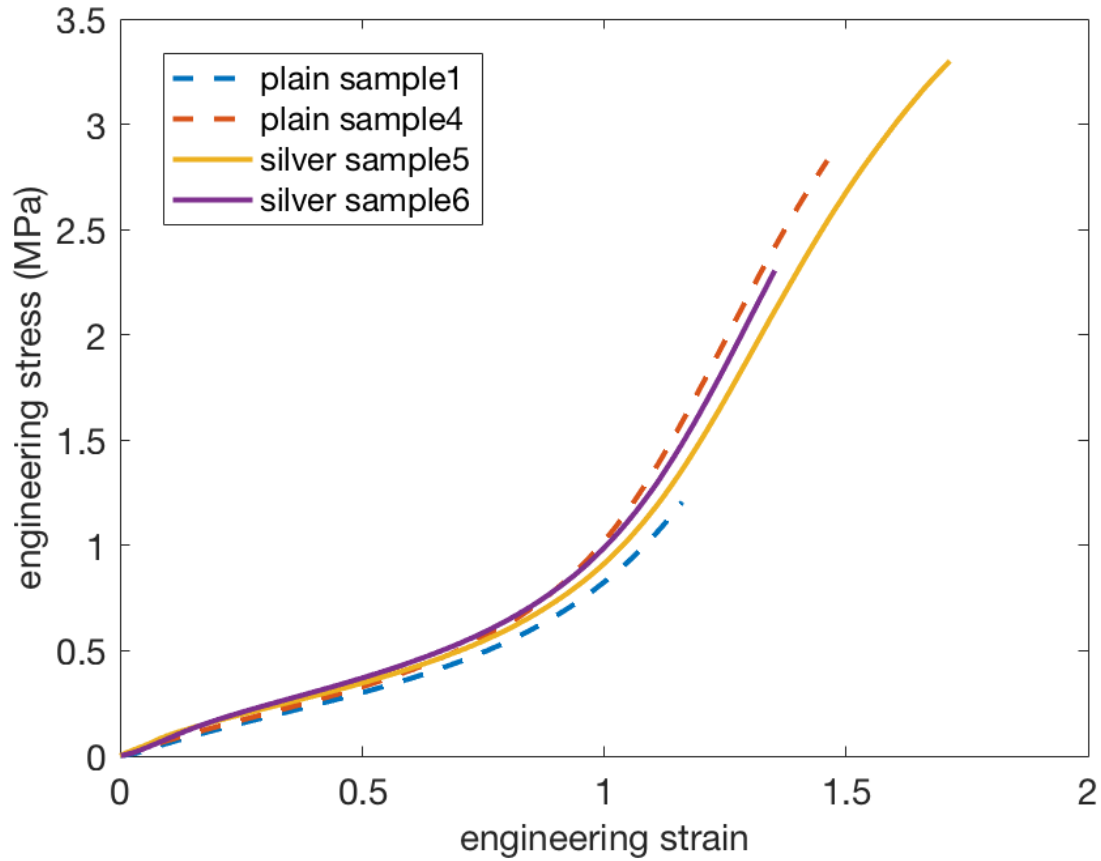


Figure 3.12: Plot of the stress-strain data measured from 4 samples of PDMS membrane. 2 samples had metalization and 2 samples did not.

presented, most are based on the fundamentals they presented. As an external force is applied to the material, the stress leads to a combination of strain and bending which can be expressed by a strain ellipsoid. For materials in the linear elasticity regime of a uniaxial deformation, Young's modulus expresses the relationship between the stress and resultant strain. However, deformations outside this regime are expressed as a strain-energy function, which relates the strain energy density of a material to the deformation gradient.

We measured samples of our membrane after fabrication to evaluate the strain-energy function. Using a TA Instruments DMA Q800, we performed uniaxial testing using a controlled force procedure to obtain the stress/strain relationship for each sample. Two samples of both metallized

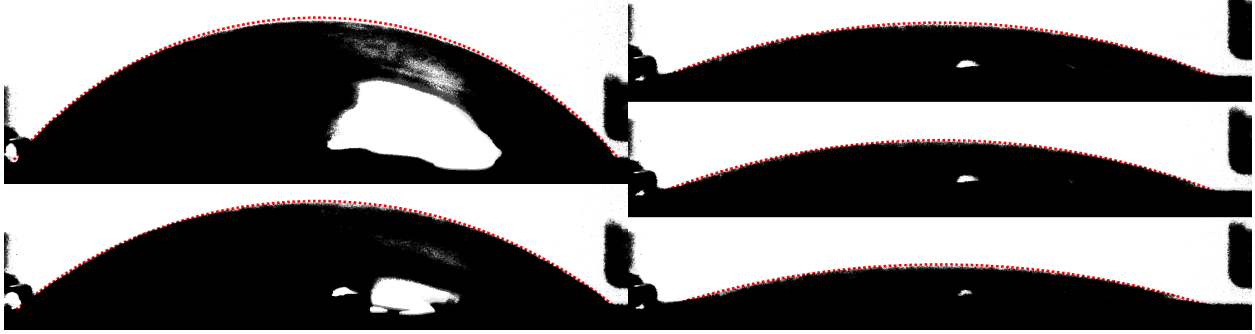


Figure 3.13: Images showing deformation shape of membrane. Left side top and bottom show extreme deformations. Right side illustrates the shapes for focusing at 800 cm (top), 50 cm (middle), and 20 cm (bottom). The red dotted lines indicate fitted circular functions illustrating how closely the membrane curvatures approximate a sphere.

and non-metallized membrane were processed. The data obtained may be visualized in Figure 3.12.

If we want to produce further enhancements in the reducing the optical aberrations of our display, a more accurate model of the membrane as it deforms is required. In pursuit of developing such a model, we captured the shape of the membrane at different curvatures as seen in Fig. 3.13. Using Canny edge detection (Canny, 1987) and filtering techniques we were able to calculate the curvature of the surface and fit polynomials by posing and solving a least squares problem.

As stated above, the equations in Section 3.3.1 are based on the assumption that as the membrane deforms, it always maintains a spherical curvature. The deformed membrane cross-sections captured in Figure 3.13 and the indicated circle functions illustrate that the membrane, while approximating a sphere, is not exactly spherical, which should be taken into account.

In attempting to control the curvature of the membrane, for a given display configuration with fixed object distance, Equation 3.3 clearly indicates there are two free variables: the current distance of deformation, which is controlled at run-time by the dynamic pressure system, and the size of the aperture, which is fixed at design-time.

When designing a deformable beamsplitter housing for a specific prototype configuration, this single design-time free parameter must be appropriately determined. Design considerations include real-world space constraints in having two deformable beamsplitters side-by-side for

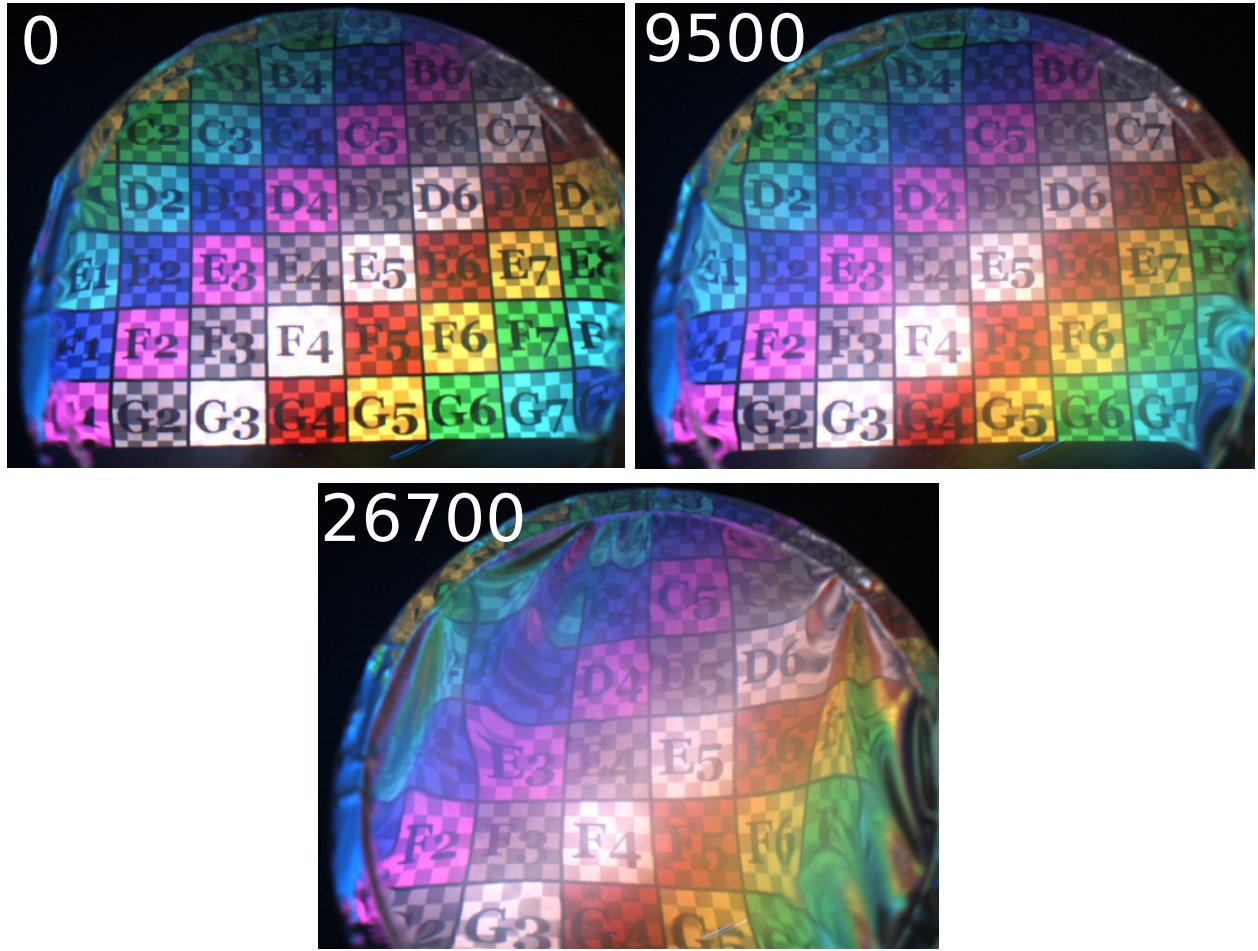


Figure 3.14: Photographs showing the result of our deformation test to estimate usability over lifetime for our in-house built deformable membrane mirror. We iterate on stretching the deformable membrane mirror back and forth with 10 times larger deformations than we use during operation. Each photograph shows a counter that represents number of iterations.

stereoscopic displays, desired focal range, focus resolution, speed of focus change, and any corrections for off-axis optics. An off-axis curved mirror will exhibit an astigmatic aberration which ideally should be minimized. One method for reducing the astigmatism is to design an elliptical aperture such that when the membrane is deformed, the closely approximated toric section reduces the astigmatism. However as the curvature of the membrane changes, a set of toric sections is produced, each with different ratios between the major and minor axes focal points, meaning different amounts of astigmatic correction are required. Thus, the aperture cross-section with the least aggregate amount of astigmatism across all toric sections should be chosen.

**Durability** Using an early prototype of the housing, we conducted a deformation test for our deformable membrane mirror as shown in Figure 3.14. During our deformation tests, we stressed the membrane to deformations that are 10 times larger than the deformations that we have during operation. Large ripples at the edge of the deformable membrane are believed to be caused by a weak attachment to the housing wearing out after 26700 iterations, which we solved in later iterations of the housing with a more secure attachment. Hazing in the images is believed to be caused by a change in surface structure after many iterations. Our deformation test was conducted over a 30 h time frame. As our membrane underwent strains far greater than during normal operation without failing, we can conclude that our deformable membrane beamsplitter and pneumatics control mechanism are suitable for long-term usage.

### 3.4 Customization for Near-Eye Displays

The goal of a varifocal see-through NED is to place a virtual image at a variable focal distance from a human subject's eyes. We approach the problem of designing optics for an OST NED with a layout shown in Figure 3.2. A display located above a user's eye is reflected from a deformable membrane mirror towards the user. Assuming a mechanism causing a uniform deflection of the membrane, the deformable membrane mirror approximates a spherical concave reflective surface, defined as

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2, \quad (3.4)$$

where  $(x, y, z)$  defines the points on the sphere surface,  $M_c = (x_0, y_0, z_0)$  defines the coordinates of the sphere center, and  $r$  defines the radius of the curvature. As a result of the deflection, the user perceives the virtual image at different focal distances that depend on the mirror curvature.

We start our design (Figure 3.2) by placing a flat membrane in front of an eye with a certain eye relief  $d_{eye\ relief}$  and aperture size  $d_{aperture}$ . We tilt the membrane with an angle  $\beta$  around the X axis. We then place the display at a distance  $d_{display}$  from the membrane, and tilt it with an

angle  $\alpha$ . Desired eye-box size,  $d_{eye\ box}$ , and  $d_{aperture}, d_{eye\ relief}, \alpha, \beta, d_{display}$  are parameters of our design.

### 3.4.1 Ray tracing model

We approached the problem of calculating the required mirror curvatures for a given configuration through a 3D ray tracing model (Spencer and Murty, 1962). The objective of our ray tracing model is to find a good mirror curvature that creates the smallest resolvable spot size. The first step of our model defines sample points  $p_e$  inside a given  $d_{eye\ box}$ . In our ray tracing routine, all sample points from a given  $d_{eye\ box}$  collectively represent a forward gazing eye aperture aligned with the optical axis (Z axis). Next, we define a sample point  $p_s$  at a desired depth  $d_{virtual}$ . We choose a  $p_s$  aligned with the optical axis again. We define multiple rays from points inside an eye box  $p_e$  traveling to a sample point  $p_s$  in depth. A single one of these rays  $R_0$  is defined as

$$R_0 = \left\{ \begin{array}{l} p_e = \begin{bmatrix} p_{e_x} \\ p_{e_y} \\ p_{e_z} \end{bmatrix}, \\ a_{es} = \frac{1}{d_{es}} \begin{bmatrix} p_{s_x} - p_{e_x} \\ p_{s_y} - p_{e_y} \\ p_{s_z} - p_{e_z} \end{bmatrix}, \end{array} \right. \quad (3.5)$$

where  $p_e$  indicates a starting point,  $a_{es}$  indicates direction cosines of the ray, and  $d_{es}$  indicates the distance between  $p_e$  and  $p_s$ . We trace  $R_0$  from pupil plane to deformable membrane mirror. Note that Figure 3.2 shows two anchor points for the deformable membrane mirror. Any given sphere that has such anchor points at both axis (X and Y) by definition is on the line that is perpendicular to the flat mirror surface, and crosses the center of the flat mirror surface. Assuming a configuration as in Figure 3.2, such a line can be defined as  $z = -\tan(\beta)(y + d_{eye\ relief})$ , leading to  $M_c = (0, y, z)$ . The intersection point between a deformable membrane and  $R_0$  can be calculated

by finding a ray propagation distance  $d_0$  that satisfies the sphere equation on the surface of the membrane with a point  $p_{mirror} = (p_e + d_0 a_{es})$ . Thus, ray propagation distance can be calculated by finding the roots of

$$\|p_{mirror} - M_c\| = r, \quad (3.6)$$

and choosing the closest root to the plane of the deformable membrane mirror. A surface normal  $R_{n_0}$  of the deformable membrane mirror at a point can be calculated as

$$R_{n_0} = \begin{cases} n_{mirror} = p_{mirror}, \\ a_{mirror} = \frac{p_{mirror} - M_c}{\|p_{mirror} - M_c\|}. \end{cases} \quad (3.7)$$

Using  $R_{n_0}$  and  $R_0$ , we calculate the reflection as a ray  $R_1$  which can be calculated as

$$R_1 = R_0 - 2R_{n_0}(R_0 \cdot R_{n_0}). \quad (3.8)$$

To calculate the intersection of  $R_1$  with a display plane, we need to be able to calculate two things: (1) surface normal of our display plane and (2) ray propagation distance  $d_1$  from the origin of the ray to a display plane. The surface normal of our display plane  $R_{n_1}$  can be calculated as

$$R_{n_1} = \begin{cases} p_{display} = \begin{bmatrix} 0 \\ \sin(\beta) d_{display} \\ d_{eye\ relief} - \cos(\beta) d_{display} \end{bmatrix}, \\ a_{display} = \begin{bmatrix} 0 \\ \sin(\beta + \alpha) \\ \cos(\beta + \alpha) \end{bmatrix} \end{cases}. \quad (3.9)$$

Using the surface normal and a vector  $R_2$  from  $p_{mirror}$  to  $p_{display}$ ,  $d_1$  can be calculated as

$$d_1 = \frac{R_{n_1} \cdot R_2}{R_{n_1} \cdot R_1}, \quad (3.10)$$

and finally, we can find the intersection point as  $p_{final} = p_{mirror} + d_1 a_{R_1}$ . We use the intersection points to calculate the spot size, in which full width at half maximum (FWHM) size of the spot diagonal is calculated using  $FWHM = 2.355\sigma$ . Using secant method, we optimize the curvature of the deformable mirror membrane by minimizing FWHM size for a given configuration. We choose a new curvature  $r_{new}$  at each iteration as

$$r_{new} = r_{current} \left( 1 - \frac{FWHM_{current} - FWHM_{previous}}{r_{current} - r_{previous}} \right). \quad (3.11)$$

### 3.4.2 Design space

Here we explore the design space of our proposal using our ray tracing model to identify the impact of  $d_{eye\ relief}$ ,  $d_{display}$ , and  $d_{aperture}$ . First, we will analyze  $d_{aperture}$ , which is defined both in the vertical and the horizontal axis. In our designs, the aperture shapes are chosen as either circular or elliptical. Adult Americans have a mean IPD of 63 mm, and their IPDs can range between 50 mm and 75 mm (Dodgson, 2004). Thus, horizontal aperture size is dictated by IPD in the nasal direction. Maximum aperture size at a vertical axis can be of any desired size to cover a larger vertical FOV. Note that the user's eyes can be decentered with respect to the optical axis of a deformable membrane mirror; thus we choose to use  $d_{eye\ box} = 20$  mm to compensate for alignment as in the case of a conventional NED design. Designs with elliptical aperture shapes can introduce perceptible astigmatism in an optical system. Such cases can easily be corrected by placing a single axis lens in between a display and a deformable membrane mirror.

Sample designs in Figure 3.15 demonstrate our findings on the effects of  $d_{eye\ relief}$  and  $d_{aperture}$  on FOV. These results suggest that shorter  $d_{eye\ relief}$  and larger  $d_{aperture}$  promise a larger FOV.



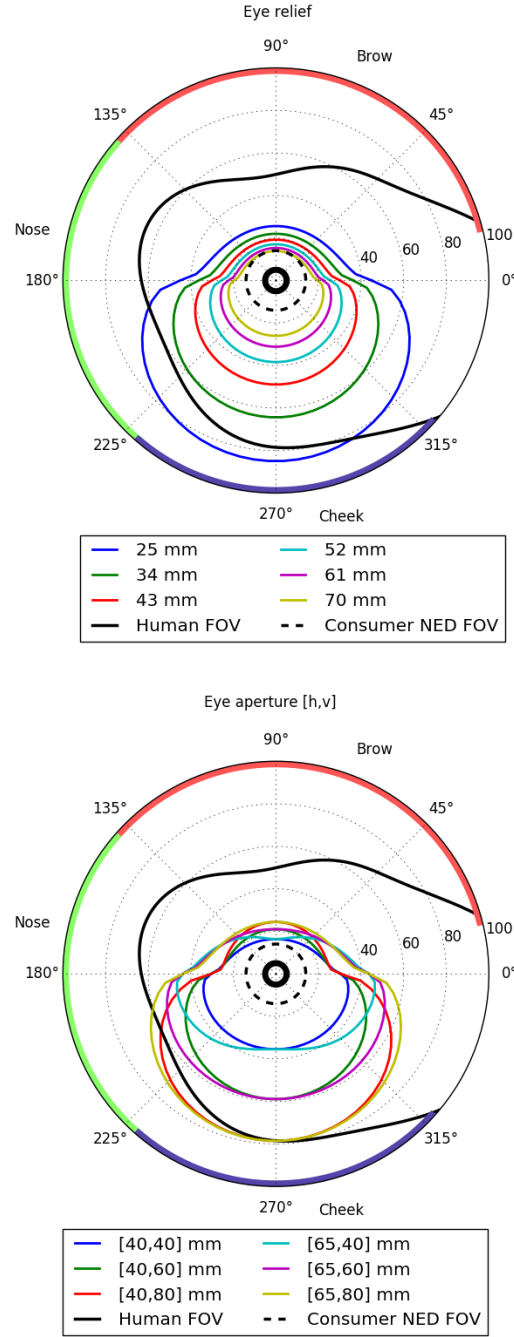


Figure 3.15: Perimetric charts in degrees showing calculated visible FOV of different sample designs for a right eye of a user while gazing statically forward. In both sketches, the solid black line represents an average FOV of a person, the solid small black circle represents foveal region, and the dashed black line represents FOV of a typical consumer level NED for augmented reality applications. Angular positions of facial features are highlighted as brow, nose, and cheek. The top figure shows variation of FOV for different values of eye relief  $d_{eye\ relief}$ . Calculations are based on a vertical aperture size  $d_{aperture_v} = 65$  mm, a horizontal aperture size  $d_{aperture_h} = 50$  mm, and deformable membrane mirror tilt  $\beta = 45^\circ$ . The bottom figure shows variation of FOV for different values  $d_{aperture_v}$ , and  $d_{aperture_h}$ . Calculated values in the bottom figure are based on  $d_{eye\ relief} = 34$  mm, and  $\beta = 45^\circ$ .

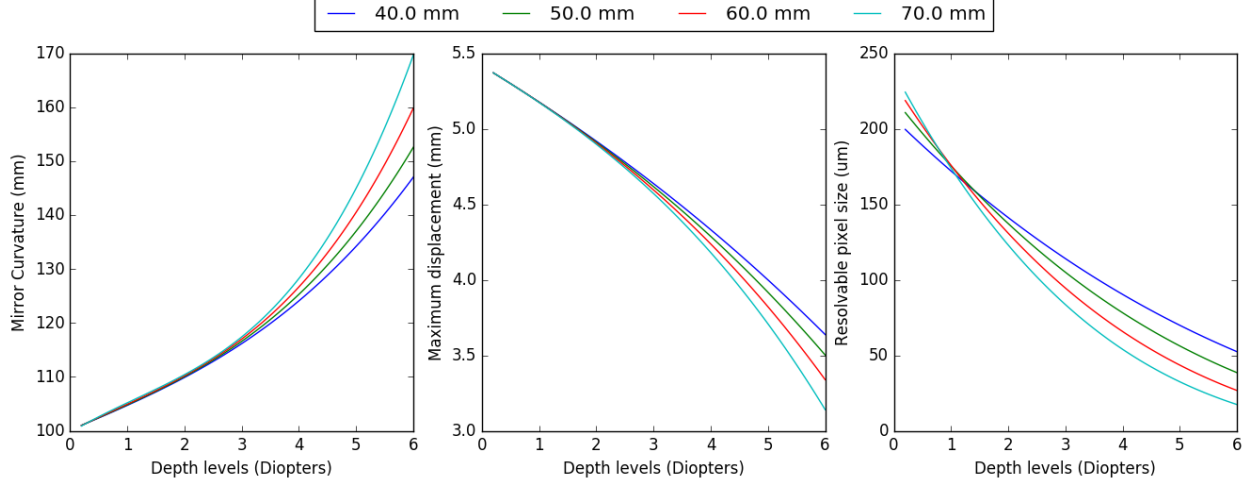


Figure 3.16: A sample design is evaluated for different eye reliefs  $d_{eye\ relief}$  with a configuration of an aperture size  $d_{aperture} = 50$  mm in horizontal axis, an aperture size  $d_{aperture} = 65$  mm in vertical axis, a mirror tilt  $\beta = 45^\circ$ , a screen tilt  $\alpha = 20^\circ$ , an eye box  $d_{eye\ box} = 20$  mm, and a screen distance  $d_{display} = 60$  mm. For all evaluations, on-axis depth fields as shown in Figure 3.2 are chosen at different depth levels. A deformable membrane mirror's curvature is calculated for different depth levels as shown on the left. The maximum amount of displacement required by each depth level is shown in the middle figure. Assuming an eye with an aperture size of 6 mm, resolvable pixel size on a screen inside the given eye box is calculated for different depth levels as shown in the figure on the right. Smaller  $d_{eye\ relief}$  benefits the design by decreasing required displacement on a membrane, however resolution improves at closer depths with a larger  $d_{eye\ relief}$ .

We would like to highlight that the majority of our designs promise a larger FOV than a typical NED for AR applications. The main limitation of our designs comes from the limited FOV generation towards the brows due to the  $\beta$  angle of the membrane mirror causing a more distant reflector in that region. Note that an asymmetrical aperture in different directions (brow, nose, cheek, peripheral), different aperture shapes (square, custom) or offsetting and angling the central axis of the membrane are possible solutions to overcome limited FOV towards the nose and the brow. However, non-elliptical designs require more complex multi-domain modeling, leading to complex surface deformations largely deviating from regular spherical or aspherical surfaces, while off-axis designs degrade the optical qualities of the reflected image by introducing optical aberrations. Increasing the aperture size will also lead to clipping the reflections of the display, particularly in the bottom region which reflects the portion of the display that abuts the brow.

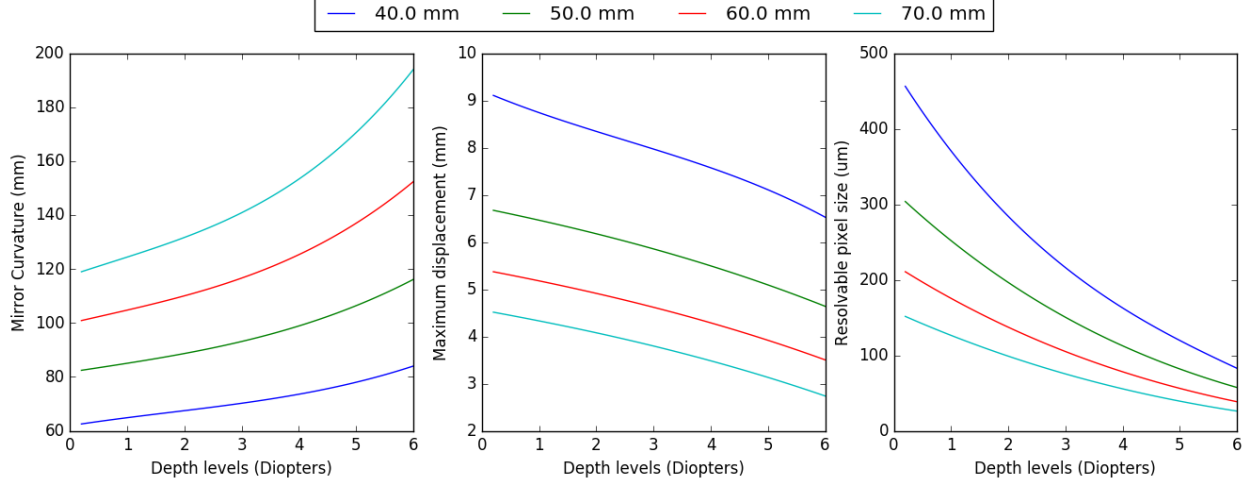


Figure 3.17: A sample design is evaluated for different display distances  $d_{display}$  with a configuration of an aperture size  $d_{aperture} = 50$  mm in horizontal axis, an aperture size  $d_{aperture} = 65$  mm in vertical axis, a mirror tilt  $\beta = 45^\circ$ , a screen tilt  $\alpha = 20^\circ$ , an eye box  $d_{eye\ box} = 20$  mm, and an eye relief  $d_{eye\ relief} = 50$  mm. For all evaluations, on-axis depth fields as shown in Figure 3.2 are chosen at different depth levels. A deformable membrane mirror's curvature is calculated for different depth levels as shown on the left. The maximum amount of displacement required by each depth level is shown in the middle figure. Assuming an eye with an aperture size of 6 mm, resolvable pixel size on a screen inside the given eye box is calculated for different depth levels as in the figure on the right.

We propose a pneumatic system to control the deformations of the membrane mirror. Understanding the required curvature values and maximum displacement for a deformable membrane mirror lets us identify the speed and volume of air movement that dictated the requirements for the pneumatic hardware. We explore the impact of different  $d_{eye\ relief}$  and  $d_{aperture}$  on curvature, displacement, and resolution by ray tracing to simulate the characteristics of different points in depth aligned with the optical axis. Our ray tracing model suggests that different  $d_{eye\ relief}$  leads to different  $M_c$ , and  $r$  configurations meaning the deformable membrane displaces different amounts with respect to the flat mirror case. We show the effect of  $d_{eye\ relief}$  with a sample design in Figure 3.16. Note that shorter  $d_{eye\ relief}$  requires less deformation of the deformable membrane mirror, which, as a result, requires more precise pneumatics. On the other hand, larger  $d_{eye\ relief}$  provides a smaller resolvable pixel size, leading to more resolution, but as noted above decreases the FOV. We conclude that the pixel size dictates the required  $d_{eye\ relief}$  in practical designs. We also evaluate the same sample designs for different  $d_{display}$ , as shown in Figure 3.17. This shows

that larger  $d_{display}$  increases resolution while decreasing the required amount of deformation on the membrane, but also increases the overall form-factor of the complete system while decreasing FOV.

### **3.5 Discussion**

Deformable beamsplitters, similar to the deformable membrane mirrors, provide a simple means for creating a large focal range in NEDs. They meet the 3 requirements of (1) being able to change focus quickly due to low-inertial operation, (2) are capable of providing a wide FOV, and (3) provide see-through capabilities with good transmission characteristics. The PDMS membrane beamsplitters are easy to produce and, using SLA 3D printed housings, assemble by hand. They can provide good quality reflected images with a wide focal range while being robust to wear and tear. In exploring the visual design space, we have shown the expected FOV for given optical configuration parameters, illustrating that they are better than or are competitive with the best OST displays available today.

## CHAPTER 4: NEAR-EYE DISPLAY PROTOTYPES

I demonstrate the capabilities and practical nature of a deformable beamsplitter display with two experimental OST, varifocal, NED prototypes. These prototypes employ deformable beamsplitters as the key optical element. Our display mitigates VAC by adjusting the optical depth of the virtual image to match the depth of the fixation point of the user. The optical depth is controlled by changing the curvature of a deformable beamsplitter membrane which reflects the virtual image with varying optical power, while transmitting an unchanged view of the real world.

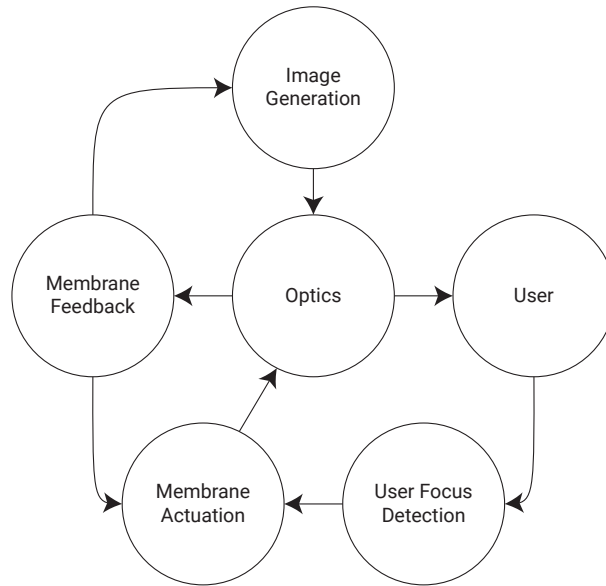


Figure 4.1: A sketch showing the overall system design and the interactions between the subsystems for all prototypes.

Each AR varifocal display prototype is made up of several subsystems which work together to make a full system. The subsystem framework used for the prototypes is explained below and their interactions are depicted in Figure 4.1.

- Image generation - This system takes information about the scene and current focal state of the user and, through rendering and post-rendering warp and via a display module, generates the appropriate image for the current frame.
- Optics - This system conveys the lighted image from the image generator to the user's eyes.
- Membrane Actuation - This system controls the curvature of the deformable membrane beamsplitter by taking input about the current focal state of the user's eyes and the current shape of the membrane.
- Membrane Position Feedback - This system monitors the membrane and reports its current state.
- User Focus Detection - This system monitors the user's current focal state and reports it.

As I describe the prototypes in this chapter, along with detailing each of the subsystems and emphasizing the unique aspects thereof, I will describe the design challenges and the reasoning behind the choices made. I will also describe the properties and capabilities of each prototype with particular interest in the limitations.

## **4.1 First Prototype Display**

For the first prototype, our choice of design parameters was mainly constrained by the availability of off-the-shelf components and the costs of custom tooling. Overall, the near-eye optics portion of our prototype consumes a volume of  $157\text{ mm} \times 205\text{ mm} \times 254\text{ mm}$ , as shown in Figure 4.4. The image generation system, optics, membrane actuation, membrane position feedback, and user focus detection systems are presented in Sections 4.1.1 through 4.1.5. An analysis of the limitations and optical qualities of this first prototype are discussed in detail in Section 4.1.6.

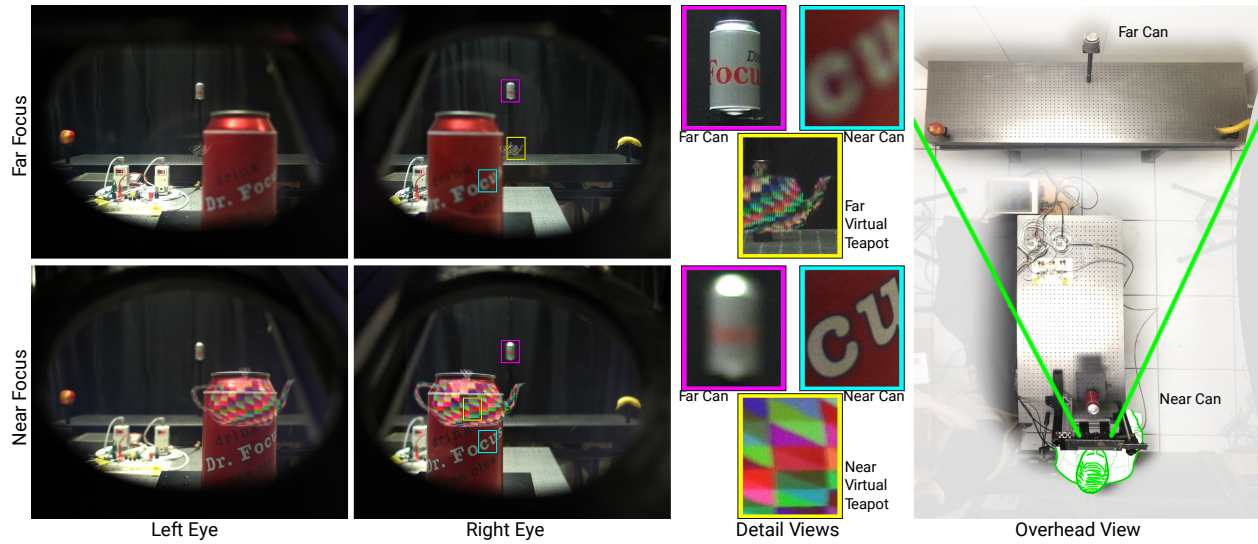


Figure 4.2: Wide FOV AR display showing virtual teapot at far and near distance together with real objects, soda cans, at near and far. Photos through first prototype display system left and right eyes with focus at far (top row), focus near (bottom row), and overhead view (right) of the system. Details from right eye views showing focus of near and far soda cans and virtual teapot (middle).

#### 4.1.1 Image Generation

We developed an in-house software to control our prototype to render images with correct distortion correction and to conduct experiments. Our software is written in the Python programming language taking advantage of GLFW<sup>1</sup> for user interface rendering, OpenCV<sup>2</sup> for image processing tasks, and Pupil-labs library<sup>3</sup> for gaze tracking tasks. Our software runs on an Intel Xeon central processing unit W5590 @ 3.33 GHz PC with two Nvidia Quadro NVS 420 graphics processing units and Linux operating system.

A one-time per housing distortion correction calibration procedure is required to correct the image distortion caused by our deformable membrane mirror at every depth. The procedure begins by selecting 8 calibration depths at 1 D spacing, then for each depth, capturing a series of gray code images with a PointGrey Flea FLEA-HICOL camera<sup>4</sup> with a Fujinon  $F1 : 1.2$  —

<sup>1</sup> <http://www.glfw.org/>

<sup>2</sup> <http://opencv.org/>

<sup>3</sup> <https://github.com/pupil-labs/pupil>

<sup>4</sup> <https://www.ptgrey.com>

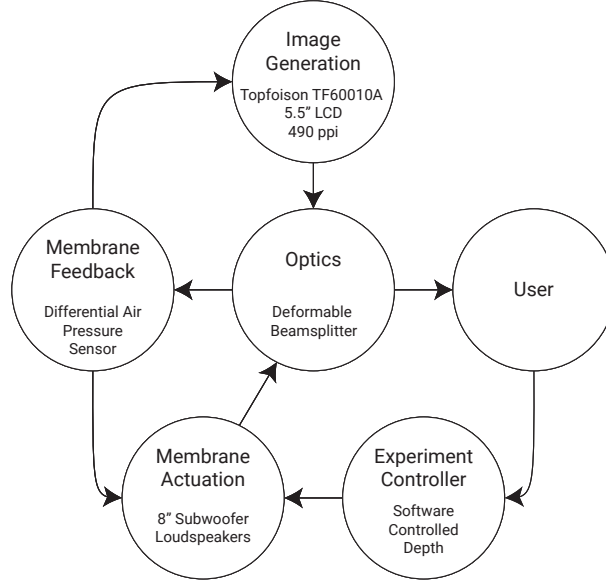


Figure 4.3: A sketch showing the system specific design and the interactions between the subsystems of first prototype.

2.8-8 mm aperture. The camera is for identification of image distortions, and not a permanent part of our system. We characterized the image distortions by using the work of Yamazaki et al. (2011) and our captured images. Once a distortion look-up table is generated for each calibration depth, we used linear interpolation to correct for image distortion for all intermediate depths in our depth range. We applied the distortion correction as an image warp shader in our rendering pipeline to present images consistent with the changing focus.

We used a Adafruit Qualia 9.7” thin film transistor (TFT) LCD<sup>5</sup> with 260 pixels per inch( ppi) for creating the optical image. The active region used per eye is 1050 px × 1260 px.

#### 4.1.2 Optics

Given the components and the evaluated sample designs, we chose the following design parameters:  $d_{eye\ relief} = 65$  mm, membrane angle  $\beta = 40^\circ$ , screen angle  $\alpha = 10^\circ$ ,  $d_{display} = 45$  mm, and  $d_{aperture} = 65.5\text{ mm} \times 80.7\text{ mm}$  (H × V). Our fully assembled prototype is shown in Figure 4.4. The FOV provided by our prototype matches our estimations computed using our ray-tracing

<sup>5</sup> <https://www.adafruit.com/product/1652>



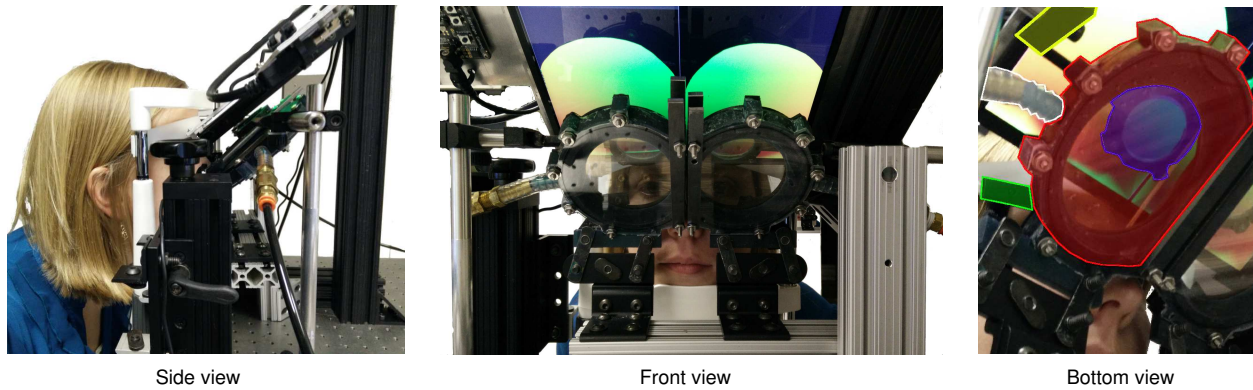


Figure 4.4: Photographs showing side, front, and bottom views of our first wide field of view varifocal near-eye display prototype for AR applications. Bottom view presents red, blue, green, yellow, and white highlighted regions, which are the deformable membrane mirror for right eye, an additional lens to overcome astigmatism in the central regions caused by elliptical deformable membrane aperture shape, a infra red camera for deformation control, a camera for gaze tracking, and a pneumatic connection to the 3D printed deformable membrane mirror housing.

model. Monocular FOV was measured as  $60^\circ$  H while the binocular FOV of our prototype was  $90^\circ \times 45^\circ$ . The membranes for this prototype were  $240\ \mu\text{m}$  thick and were fabricated by spin coating at 300 rpm. An additional 6 D cylindrical lens was placed in front of the central regions of the LCD for each eye in an effort to minimize the astigmatism caused by the off-axis optics.

### 4.1.3 Membrane Actuation

We used a vacuum pump and tank (115 torr  $\sim 15\ \text{kPa}$ ) as a vacuum source. Per eye, we had a SMC ITV2090-21N2BL5<sup>6</sup> vacuum regulator, a T-junction, and a bleed hole to create a controlled partial vacuum environment inside our vacuum housing. Our vacuum regulators were capable of regulating pressure levels in between  $-1.3\ \text{kPa}$  to  $-80\ \text{kPa}$ , and each was controlled by a Teensy 3.2 microcontroller<sup>7</sup>. Our combination of microcontrollers and vacuum regulators provided us  $\sim 60$  addressable stable depth planes ranging from 0.2 D to 7 D according to our empirical experiments. The two microcontrollers coordinated by getting signals from, and sending messages to a PC over universal serial bus (USB).

<sup>6</sup> <https://www.smc-pneumatics.com>

<sup>7</sup> <https://www.adafruit.com/product/2756>

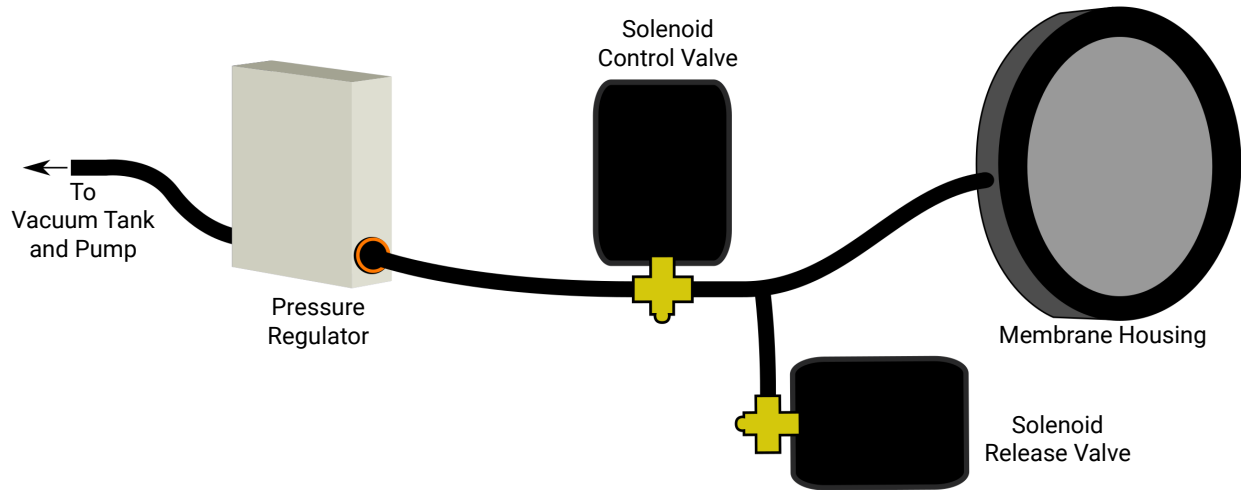


Figure 4.5: A sketch showing the vacuum system of first prototype.

#### 4.1.4 Membrane Feedback

Our control methodology for the deformations of the deformable membrane mirror was based on reflection shape detection from an infrared radiation (IR) light-emitting diode (LED) placed above each deformable membrane mirror. A Samsung Series 9 NP900X3A NP900 Webcam BA59-02904A SC-13HDL10931N camera with an Edmund Optics Optical Cast IR Longpass Filter placed in front of the lens running at 30 frames per second( fps) for each deformable membrane mirror was positioned above the deformable membrane mirror as shown in bottom view of Figure 4.4. When the system was directed to change the effective focal power, the PC electronically controlled the vacuum regulator through the microcontroller using the reflection detection from the IR cameras which acted as a feedback mechanism to form a closed-loop control mechanism.

#### 4.1.5 User Focus Detection

Our prototype used a gaze tracking Pupil-labs camera<sup>8</sup> per eye, running at 120 Hz attached via custom 3D printed mounts to determine the gaze vector for each eye. Then using the depth

---

<sup>8</sup> <https://pupil-labs.com/store/>

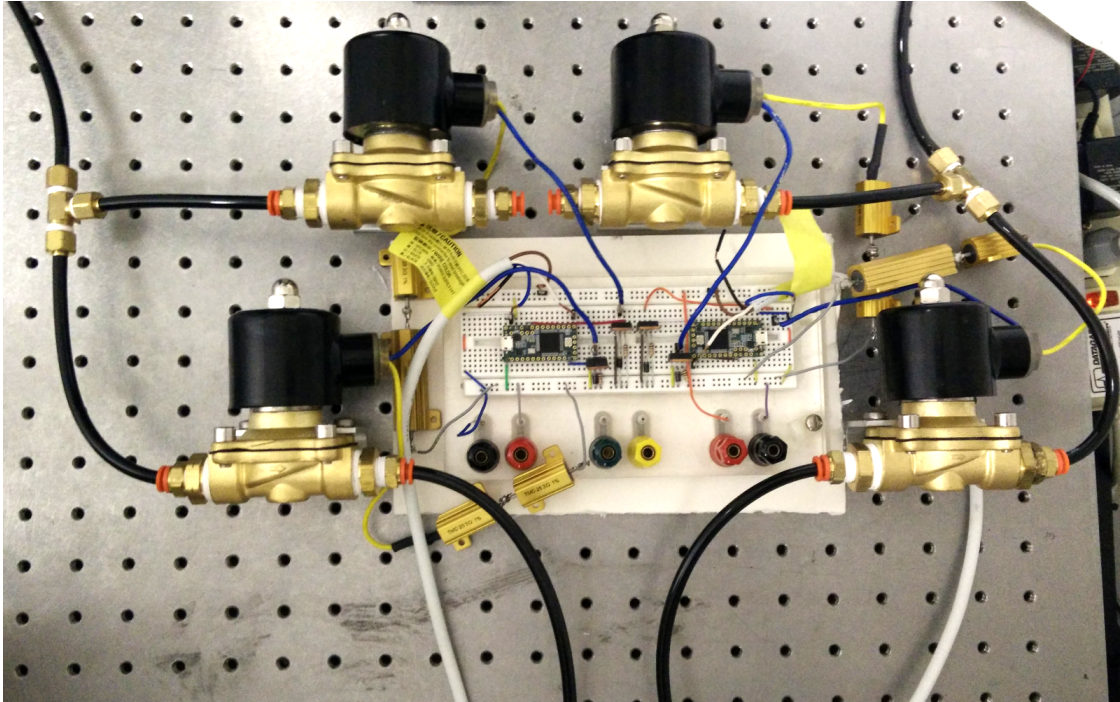


Figure 4.6: A photograph of the electronics which controlled the operation of our first prototype. Two microcontrollers, one per eye, independently drove the vacuum regulators and solenoid valves.

from vergence algorithm described in Section 5.2 we determined the depth of the point of fixation which, due to vergence-accommodation coupling, is equivalent to the user's focal depth.

#### 4.1.6 Limitations

**Pneumatics** The first prototype response time of 300 ms for switching from one extreme depth level to another was shortened by revisiting our pneumatics hardware design for our second prototype. The pneumatics of our first prototype created a low-volume, audible noise as it accommodated to different image planes, however in our second prototype we prioritized silent operation and eliminated depth-switching noise. In the second prototype the only audible noise comes from the leakage correction system, which operates infrequently. A two-compartment based design could also help avoid noise by simply pushing air from one compartment to the other.

**Form-factor** Our optical design can be less bulky in size. According to our ray tracing model, tilting the deformable membrane mirror away from the nose can shrink the size of required aper-

ture leading to a smaller overall size. Another challenge for our first prototype was shrinking the optical path in between a display and a deformable membrane mirror,  $d_{display}$ . Through our analysis, with an increasing  $d_{display}$ , we observe that display size grows, and the supported resolution increases. We addressed this challenge in the second prototype by revisiting the image generation system in our design. The higher ppi, smaller LCD panel in the second prototype was positioned closer to the forehead at smaller angle  $\alpha$ . The denser resolution offset the loss of resolution caused by a shorter optical path, while the decreased angle decreased the amount of astigmatism caused by the off-axis curved beamsplitter. These modifications on the second prototype assisted in moving towards a wearable version enabling further experimentation in AR application specific research.

**Latency** The performance of our display is affected by the system latency. End-to-end latency is a combination of many elements. Its first source is the eye tracking system, which can be changed by using an alternative solution. Good quality eye tracking systems can achieve latency as low as 50 ms (Guenter et al., 2012).

The second source of latency is the membrane deformation latency. The change of its shape for the most extreme scenario can take up to 300 ms. Again, these limitations may remain unnoticed due to the characteristic of the eye accommodation process which also exhibits large delays as explained in Section 2.2.

The total latency of our system remains below the delays of the eye accommodation process of 500 ms to 1300 ms and may be sufficiently low for AR applications. This is supported by results of our subjective experiments. Further experiments regarding reducing the latency required for perceiving stimuli are described in Chapter 6.

The operational velocity of our membrane amounts to 16.6 D/s, which might be below the peak velocity for disaccommodation in extreme depth changes. Since our membrane deformation is initiated during the period of eye accommodation latency and its maximum duration is less than 300 ms, we expect that the whole process is completed well before such extreme lens accommodation velocities are reached.

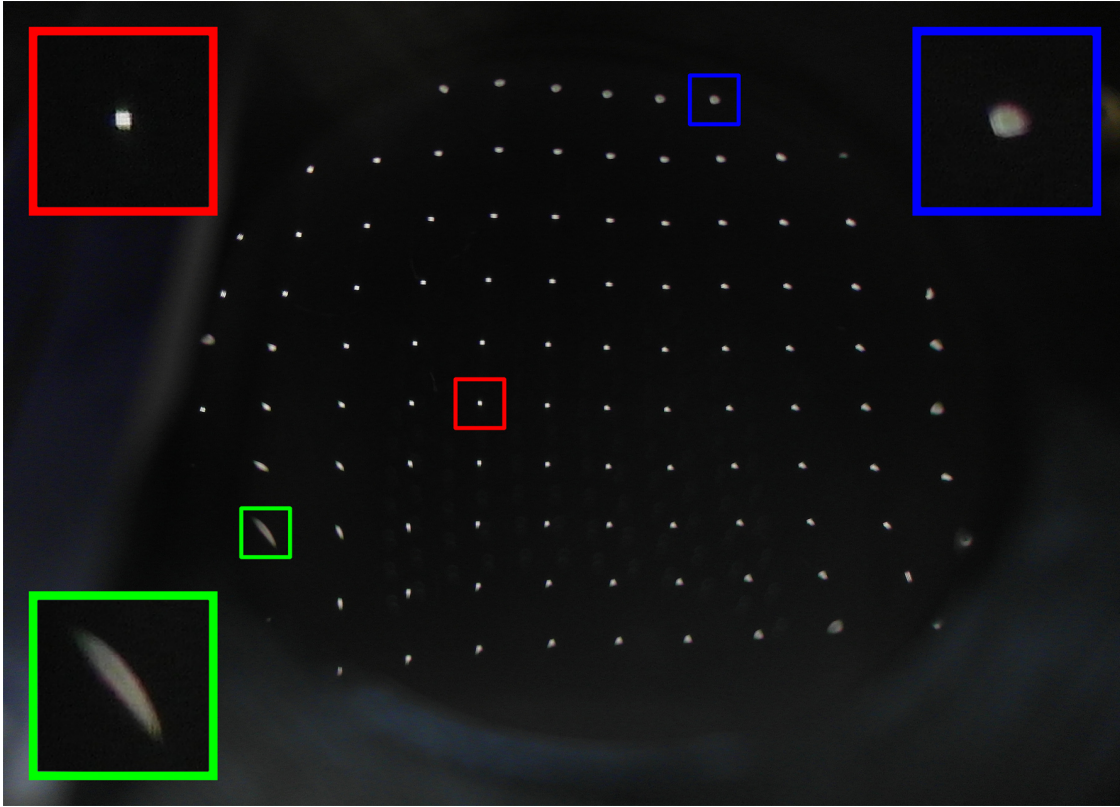


Figure 4.7: A view approximating the point spread function across the membrane of our first prototype. Squares of 3x3 pixels are illuminated in a grid pattern to show the graceful degradation of focus across the membrane. Central region shows minimal point spread (red inset), while periphery shows a much larger point spread (blue inset). Severe defocus in lower left region (green inset) is caused by inadequate tension on membrane when closing and securing the housing.

In contrast to the discussed findings from the perception literature, the accommodation state of our prototypes are changed by the systems in response to the detected change in user vergence. While the literature indicates that for stimulus-driven gaze changes vergence occurs faster than accommodation, to the best of our knowledge, no studies have been performed under the conditions created by our display. The delayed focal information presented by our display may have an effect on user accommodation response.

**Consistency of focal properties** Our display does not guarantee the same focus properties across the entire FOV as this would require a more challenging membrane design by optimizing its mechanical properties. Instead, our system provides a correct focal cue in the central field and has a gradual degradation towards the peripheral field as seen in Figure 4.7. This aligns with



the limitation of the human visual system regarding the eye sensitivity in the peripheral regions as described in Section 2.2.

The increase of DOF with eccentricity reduces requirements imposed on the membrane design in our display, as relatively high defocus blur can be tolerated outside the central foveal region without causing any perceivable degradation of the image quality. Using the eye tracking system, we are able to provide a precise focus in the fovea region, while the precision of membrane shaping outside such region can be relaxed. This greatly simplifies maintaining the high visual quality over a wide FOV. It is left as future work how much improvement is perceived when there is correct focus across the entire membrane.

**Depth of field** Our display is capable of displaying only a single depth at a time, which leads to incorrect views for virtual content at different depths. A simple solution to this would be to apply a defocus kernel approximating the eye’s point spread function to the virtual image according to the depth of the virtual objects. Due to the potential of rendered blur not being equivalent to optical blur, we did not implement this solution for the first prototype.

**Occlusion support** The works of Kiyokawa et al. (2003), Gao et al. (2012), and Hamasaki and Itoh (2019) describe occlusion capable NEDs, and introduce application spaces that require occlusion support. Our system does not attempt to support occlusion. We leave this challenge as a future work.

## 4.2 Second Prototype Display

In comparison to our first prototype described in Section 4.1, the goals of our new display were threefold: (1) eliminate the extra optical element, (2) improve the image quality, and (3) decrease the overall display dimensions. Our previous prototype, which was reliant on an additional corrective lens, was large enough that it could not easily be head-mounted and suffered from a strong astigmatism due to its off-axis optical design. In addition to improvements in optical qual-

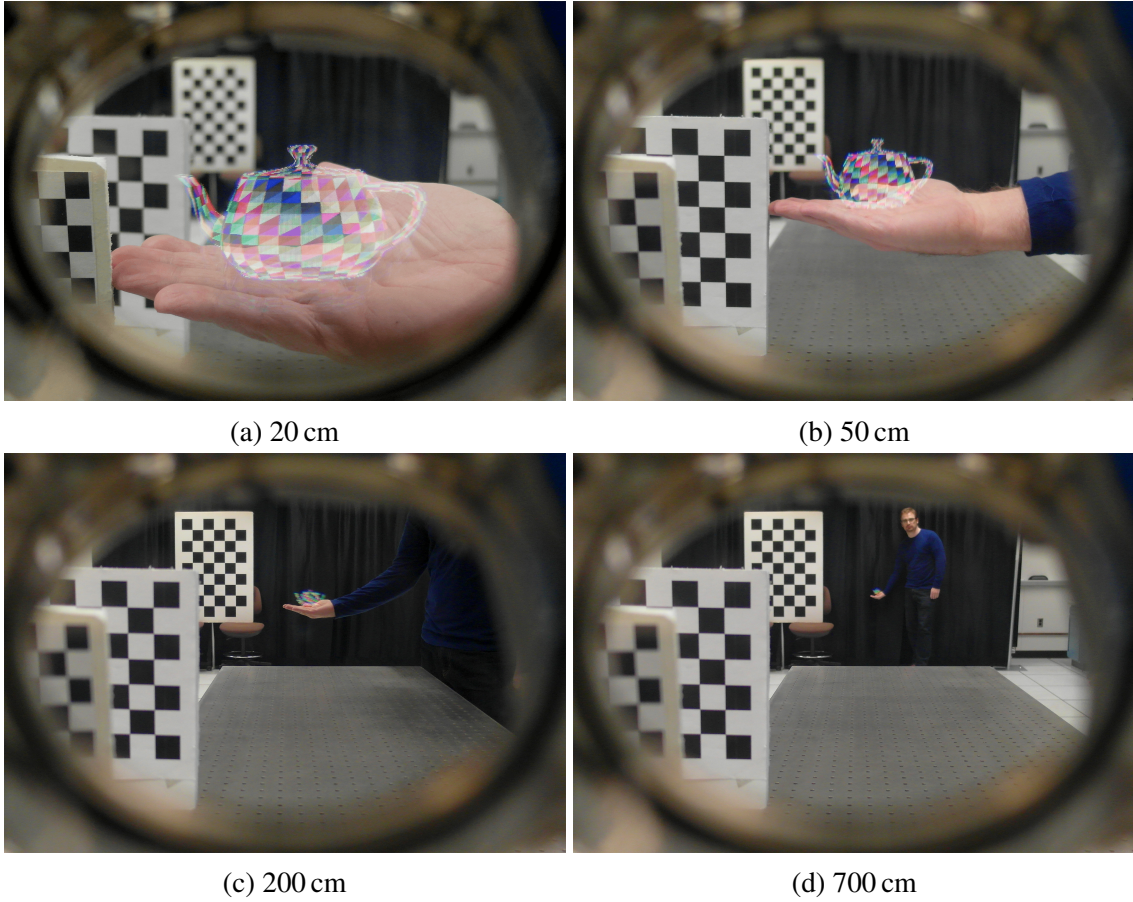


Figure 4.8: First display prototype displaying virtual teapots in hand at several depths.

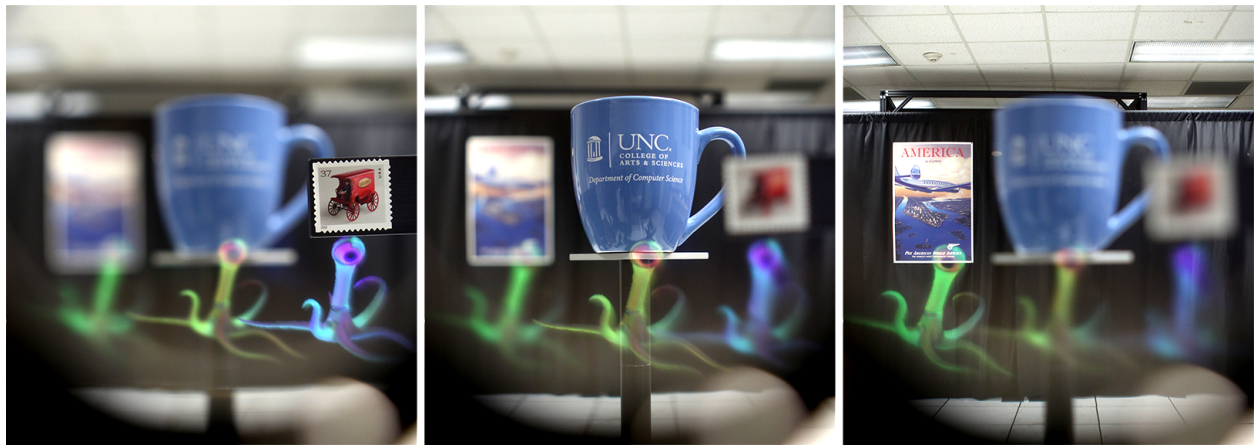


Figure 4.9: Images looking through second display prototype showing three real world objects (top) and three virtual aliens (bottom) at three different depths: 20 cm (stamp and blue alien), 50 cm (mug and yellow alien), 800 cm (poster and green alien). Virtual objects not at current focal depth are computationally blurred.



Figure 4.10: Side and front view of second prototype with illustrated user to show the form-factor. The LCD panel of this prototype is angled less with respect to the deformable beamsplitter housing than the first prototype. This improved angle enables more light to enter the optical system and improves the focal characteristics of the display.

ity and form-factor, our second prototype fully implements the deformable beamsplitter design in its simplicity: a single optical element to control the focus of the virtual images.

Overall, the head-mounted portion of our prototype, as shown in Figure 4.10, consumes a much smaller volume of  $5.5\text{ cm} \times 12.5\text{ cm} \times 15.2\text{ cm}$ , and is much lighter compared to our previous prototype. The LCD panel with cables and housing are 132 g, while each membrane, housing, and tube weigh 81 g, with a total mass of 452 g including the unoptimized mounting hardware. Improvements in optical qualities are discussed in detail in section 4.2.6. Details of our second prototype follow.

### 4.2.1 Image Generation

To generate the images for both eyes, we used a single LCD panel Topfoison TF60010A-V0  $1440\text{ px} \times 2560\text{ px}$  5.98" TFT LCD which has 490 ppi.

As before, our images are rendered in real time using an open source Python and OpenGL library dGraph<sup>9</sup>. Due to the nature of varifocal displays, as the virtual image matches the optical depth of the user's current gaze, objects that are virtually at different depths also appear

<sup>9</sup> <https://github.com/qenops/dGraph>



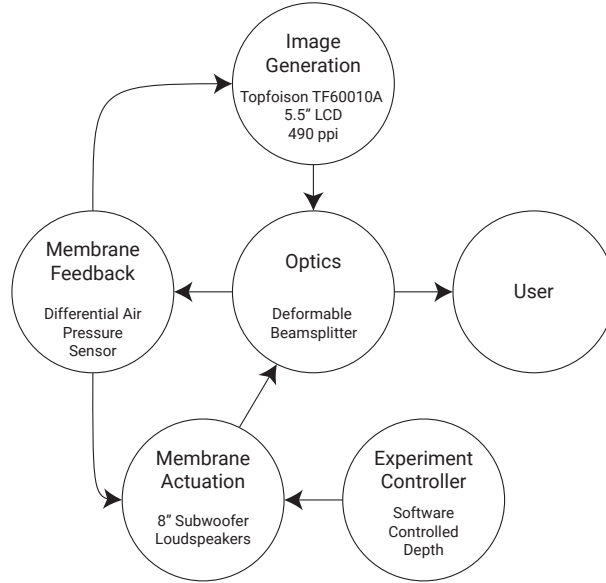


Figure 4.11: A sketch showing the system specific design and the interactions between the subsystems of second prototype. The user focus detection system was replaced with a software controller.

sharp. To solve this problem, in our second prototype, rendered blur is added in a convolution pixel shader by computing the circle of confusion between the focal depth and virtual object depth. This is then converted into pixel space by generating a 2D top hat kernel. The results can be seen in Figure 4.12. It should be noted that while the convolution kernel can be a separable function, for large amounts of blur, a fast fourier transform-based convolution would be faster. Additionally, more perceptually accurate blur can be achieved by adding chromatic aberration, as seen in Cholewiak et al. (2017). Additional distortion and anomalous perspective correction are performed in pixel shaders using lookup tables.

#### 4.2.2 Optics

The new parameters for our second prototype as compared to our first prototype are presented in Table 4.3. By simultaneously reducing the screen tilt to  $0^\circ$  and decreasing the membrane tilt to  $20^\circ$ , we reduced the astigmatism and field curvature aberrations and increased the brightness of the virtual image. A smaller, higher pixel density display is used to increase the angular resolution while shrinking the form-factor. With the placement of the lens and display fixed, an analysis

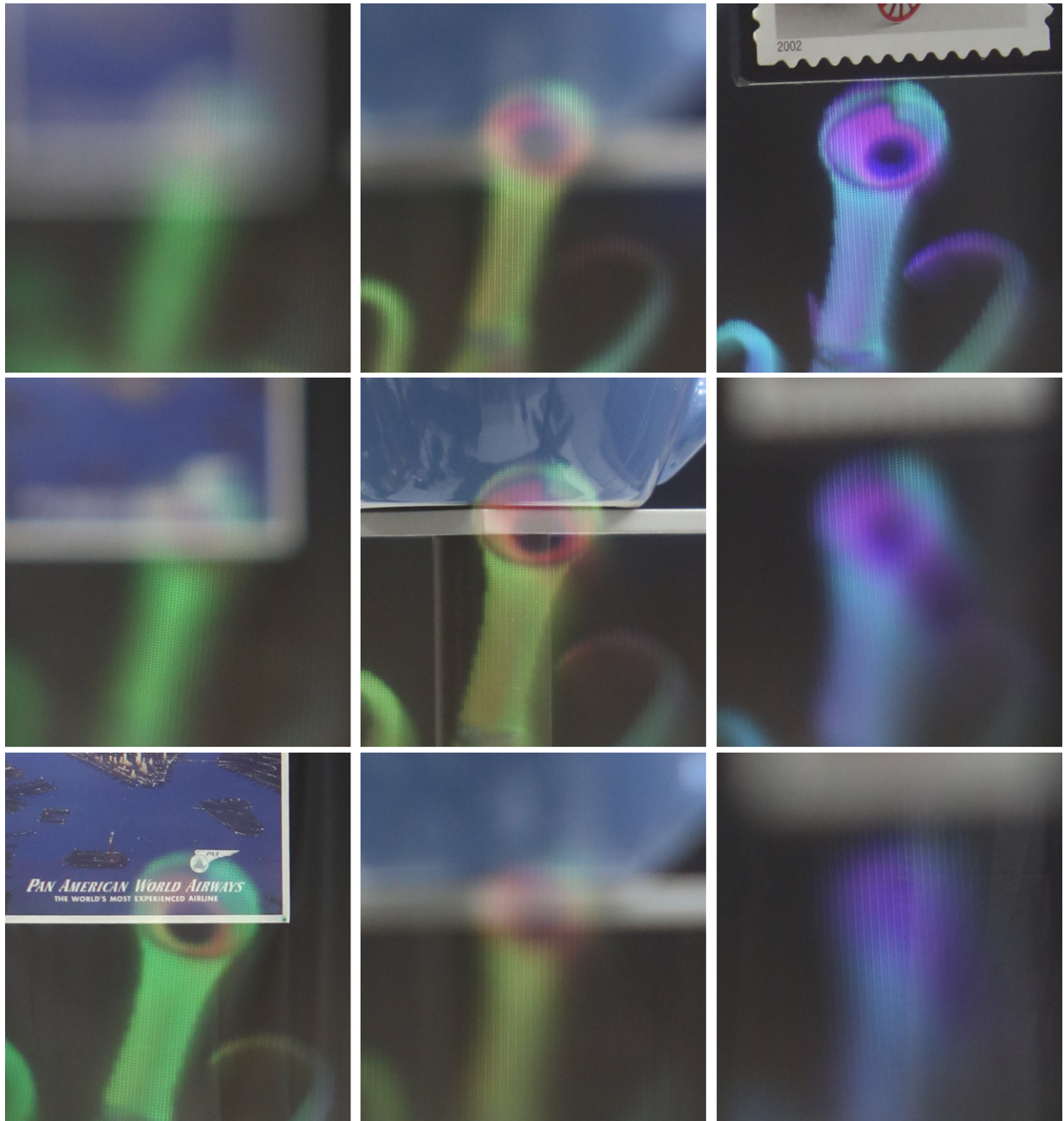


Figure 4.12: Close up view of images presented in Fig. 4.9. The near (right), medium (middle), and far (left) real and virtual objects are shown with the display and camera focus set to near (top), medium (middle), and far (bottom). Computational blur is applied to virtual objects not at current depth.

of the membrane housing aperture was done. By approximating the deforming membrane with a series of toroidal sections, an optimization reducing the spot size reflected off the torus for seven focal depths from 10 cm to optical infinity was performed in Zemax OpticStudio, giving the ratio between the minor and major axis of the torus at each depth. By weighting the greater depths more, a ratio of .8733 minor to major axis was determined. Based on the fabrication and mechanical constraints of the physical system, we iterated toward an aperture size meeting that ratio, leading to an aperture of 57.25 mm×50 mm. Keeping in mind a diverse user group, an adjustable IPD was included in the design allowing native IPDs from 60 mm to 78 mm. The lower end of the human IPD range from 50 mm to 60 mm (Dodgson, 2004) is not accommodated directly, but is fully covered by our large eye box, as reported in Section 4.2.6.

Our deformable membranes and housing for each eye are fabricated and assembled in-house using the methods detailed in Section 3.2. This prototype used 120  $\mu$ m membranes fabricated by spin coating at 600 rpm.

#### **4.2.3 Membrane Actuation**

Eliminating the vacuum pump and pressure regulators was achieved by using a Pyle PLMRW8 8" 400 W 4  $\Omega$  Marine Subwoofer to modulate the air pressure in the membrane housing for each eye.

A single Arduino Teensy 3.6 microcontroller, which uses a software proportional integral derivative (PID) controller to hold the membrane at the target depth based on the sensory inputs, drives the vacuum system as directed by the PC over USB. The speakers are driven through use of a WGCD L298N Dual H Bridge direct current (DC) Stepper Module with a 12 V 5 A DC power supply.

Leak correction and re-pressurizing the system is provided by a PeterPaul 72B11DGM 12/DC solenoid valve venting to atmosphere. The operation of the subwoofer controlled membrane is reliant upon operating within a closed chamber. While every effort has been made to ensure an air-tight system, during operation some air leakage does occur. Re-pressurization for leakage cor-

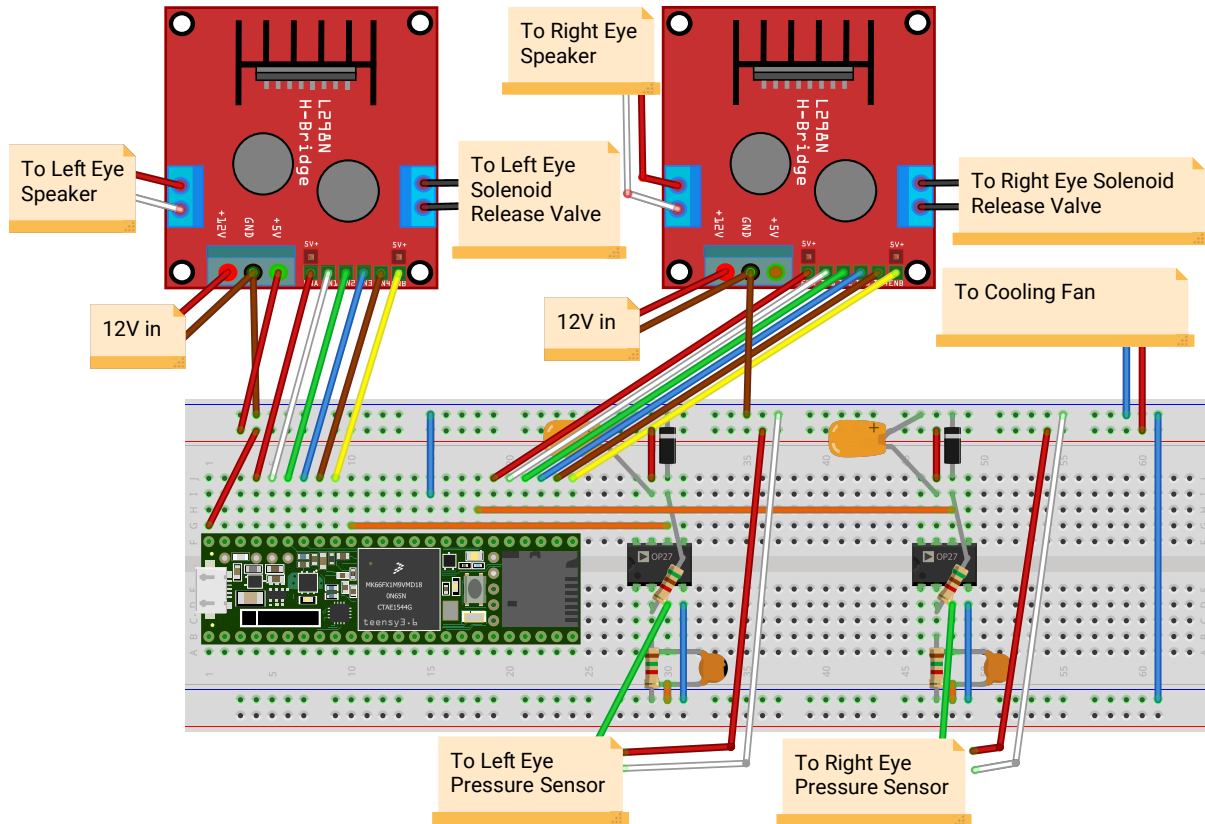


Figure 4.13: Circuit design for controlling the membrane actuation and membrane feedback systems of second prototype.

rection during continuous operation typically occurs about once every 20 min. All pressure modules are connected with SMC Pneumatics 0.25" OD Tubing, one touch fittings, and T-junctions.

#### 4.2.4 Membrane Feedback

A Motorola MPX5010DP pressure sensor provides feedback on the current pressure differential between the ambient atmosphere and inside our membrane housing, thus our system no longer uses cameras in the pressure control subsystem.

#### 4.2.5 User Focus Detection

From our experience in using image feature based gaze trackers to calculate depth from vergence in prototype 1, we determined via the process explained in Chapter 5 that currently avail-

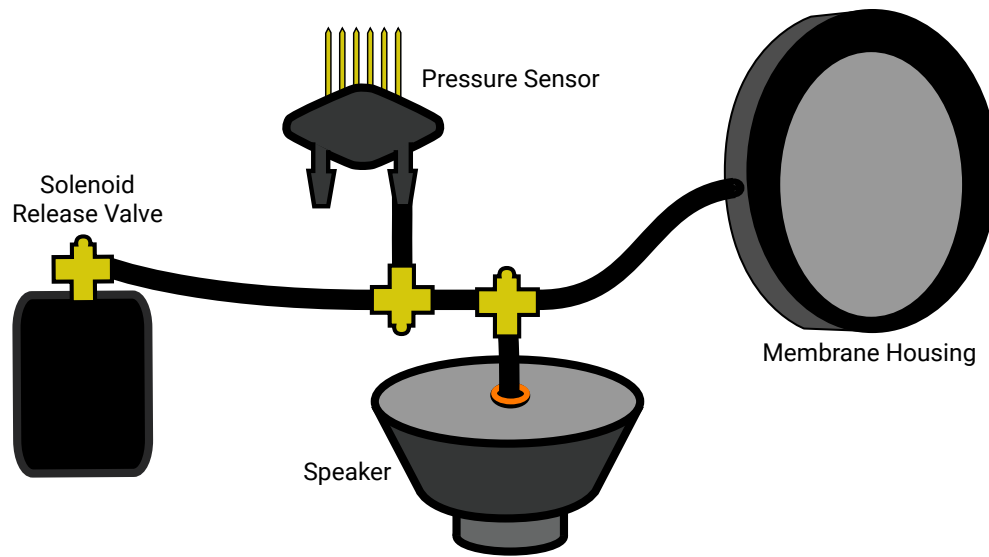


Figure 4.14: A sketch showing the vacuum system of second prototype.

able commercial gaze tracking solutions are unable to provide the accuracy required to calculate the user's fixation depth to a useful degree of accuracy. Therefore we did not employ a user focus detection system in our second prototype, instead opting for a software controller to set the depth of the display. The software controller operates by executing a limited number of hand-crafted scenarios and carefully-designed experiments which both guides the user's gaze and sets the depth according to the current state of the program.

#### 4.2.6 Optical Quality Analysis

In this section we evaluate several attributes of our display directly comparing them to studies of human performance.

**Field of View** We measured our field of view by placing a Samsung EX2F camera in the center of the eye box and capturing a checkerboard of known size at a known distance. The view from the camera was aligned to the view of a user via real-world object alignment to ensure accurate measurement. Using OpenCV we perform an image undistort to generate a distortion-free image on which we can use the size of the checkerboard to give us the angular resolution of the undistorted image. Finding the FOV is then just the process of measuring the width and height of the

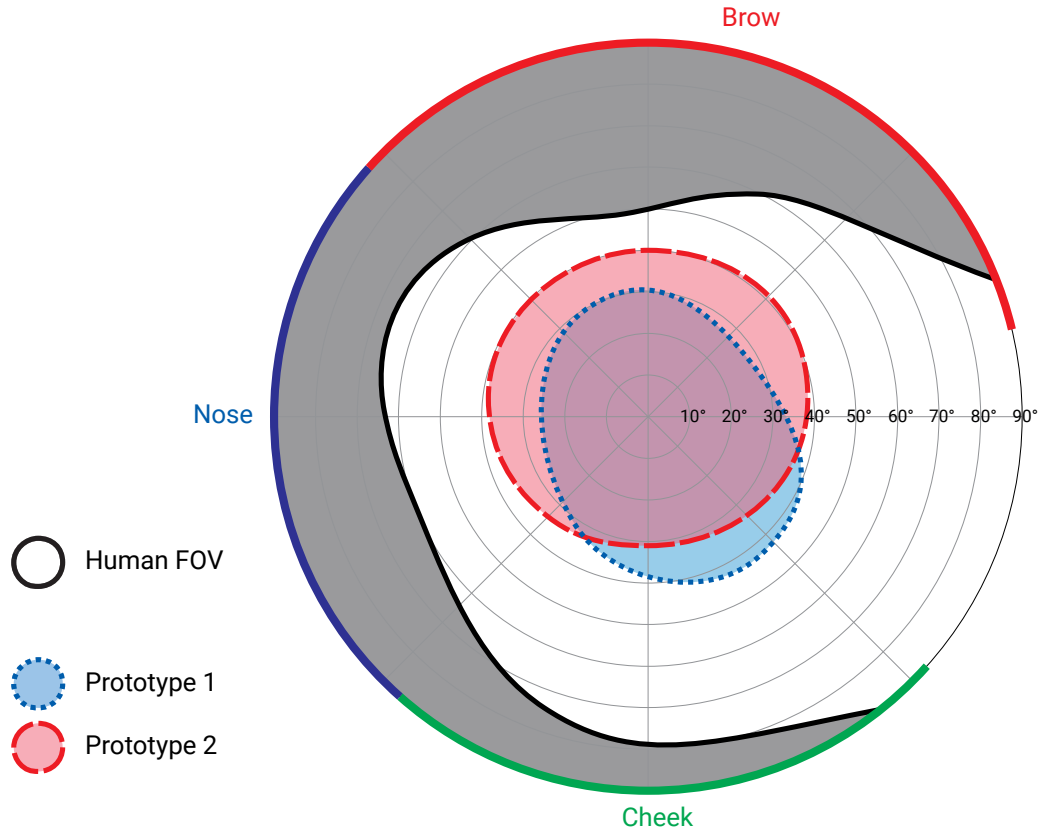


Figure 4.15: The field of view of our two prototype displays compared with an average field of view of a human right eye.

boundaries of a displayed solid white image. Our second display prototype exhibits a  $75.5^\circ$  horizontal and  $70.6^\circ$  vertical field of view; a more accurate representation can be seen in Figure 4.15.

**Focal Range** The focal range can be measured by determining the distance to the nearest and furthest virtual images the display is capable of presenting by using Equation 3.3 presented in Section 3.3.1. While our display is capable of deforming the membrane in a convex manner, it is not done during typical use, so we measure the virtual image distance for a flat membrane as the nearest. Using Equation 3.3, the membrane deflection required for presenting images at optical infinity is determined. With our prototype being capable of going beyond that deflection, infinity is used as the farthest depth. With a range greater than 15 D, our display is capable of matching the highest focal ranges of 10 year olds, with a focus between 68 mm and optical infinity (represented by 8 m in all measurements below).

**Focal Latency** As explained in Section 2.2, the lens in human eyes has a finite accommodation response with several defining characteristics: a latent period of around 300 ms, a main focus adjustment period determined by the distance of focus change, and a settling period where minor corrections are made until a state of micro fluctuations near the target is reached (Schor and Bharadwaj, 2005).

Using a GoPro Hero 4 camera at 240 fps to record the response of the membrane, we measured our prototype's performance. We visually indicated the initial signal by changing the image on the display and waiting for frame buffer swap before sending the new depth signal to the microcontroller. In all cases, our display exhibited an initial latent period less than our sampling period of 4.16 ms. Our display also exhibited an initial main focus adjustment period followed by a settling period similar to an eye. The main focus adjustment period of our prototype demonstrated a mean velocity of 55 D/s. Mean time for the initial adjustment was 139.5 ms with a maximum of 200 ms. The settling period exhibited several cycles of overshoot due to the method of PID control used, but came to rest in a mean of 201 ms and a maximum of 237.5 ms. Total adjustment times had a mean of 340 ms and a maximum of 438 ms. The long settling times indicate that improvements can be made either by tuning the PID parameters or with a better control algorithm.

**Angular Resolution** To determine the spatial and angular resolution limits of our display, we evaluate the MTF of our latest prototype at various depth levels. Our measurements are based on the International Standards Organization 12233 slanted-edge MTF method (Burns, 2000). Figure 4.16 shows the MTF of our prototype at distances 10 cm, 20 cm, 33 cm, 50 cm, 100 cm and 800 cm, all captured with a Samsung EX2F camera placed 40 mm behind the display. The camera aperture was set to f/3.9 with exposure times of 0.1 s.

First, an image is captured of a high resolution printed checkerboard pattern of known distance and size, which is used to measure the angles resolved per camera pixel. Then, a slanted-edge image is captured by the camera through our prototype, from which a specific region of interest near the center of the field of view is used to measure the MTF of the display. Low fre-

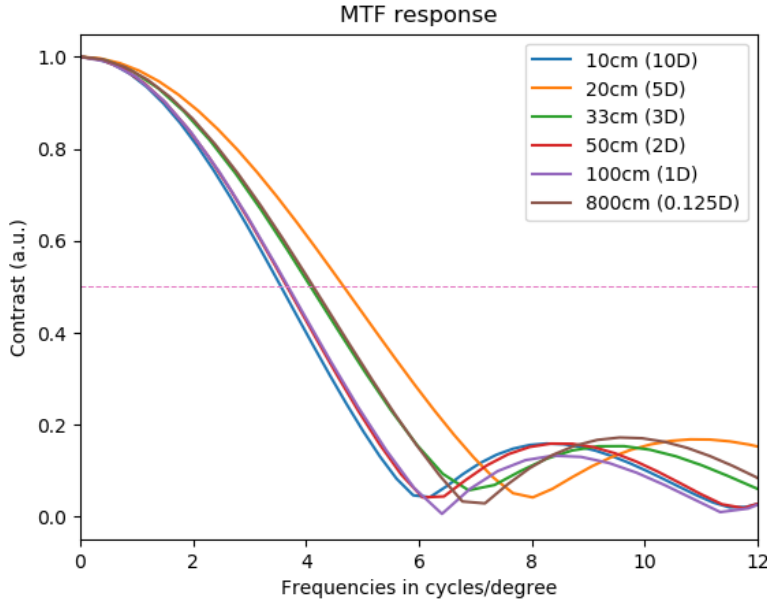


Figure 4.16: MTF of second prototype measured at various depth levels in accordance with the ISO 12233 slanted-edge method. Our prototype produces 4-6 cpd spatial resolutions at various depths.

quencies that add noise to the measurements are filtered by thresholding the edge-spread at 10 % and 90 % of the measured intensities. This process is repeated for all reported depths. It can be seen that the display is capable of consistently producing a spatial resolution of 4-6 cpd. The limitation of the spatial resolution of the display primarily comes from the available resolution of the LCD panel used for providing imagery to the eyes. In fact, the individual pixels of the LCD panel are discernible from the reflection off the deformable beamsplitter membrane. A two-fold increase in the resolution of the LCD panel results in a spatial resolution of about 14 cpd, which is the current state-of-the-art for commercially available VR displays. A slightly decreasing trend of MTF is seen with increasing distance of the virtual image, and this behavior is caused by the anomalous perspective of the virtual image combined with a more severe astigmatism.

**Eye box** A display must be able to generate an eye box capable of entering the pupil as the eye moves around the visual field, with some additional tolerance for imprecise alignment, adjustment while being worn, and variations of human anatomy. Most displays target a 20 mm×20 mm eye box (Cakmakci and Rolland, 2006; Tsurutani et al., 2017).



Table 4.1: Luminance values of display prototype given in candela per meter squared for different focal depths.

Depth (cm)	Luminance (cd/m <sup>2</sup> )
10	195
20	135
30	134
50	131
100	127
800	115

Table 4.2: A comparison of the subsystems for the two display prototypes including size, mobility, and speed.

	First Prototype	Second Prototype
Image Generation	9.7" LCD – 260 ppi	5.5" LCD – 490 ppi
Optics	Deformable Beamsplitter + Extra Lens	Deformable Beamsplitter
Membrane Actuation	Vacuum Pump and Tank	8" Subwoofer
Membrane Tracking	Infrared LED and Camera	Pressure Sensor
User Focus Detection	Gaze Tracking - Depth from Vergence	NONE
Head-mounted Size	157 mm × 205 mm × 254 mm	55 mm × 125 mm × 152 mm
Total Mobility	Tethered to Vacuum Pump	Mobile with Subwoofers
Speed	20 ms near-to-far; 300 ms far-to-near	200 ms both

We measured the eye box by attaching a camera to a 2 axis linear stage and evaluating the images captured. Measured with an eye relief of 40 mm from the membrane, the eye box for a 10 cm focal depth is 40 mm horizontal, 20 mm vertical. For all other depths, it is 30 mm horizontal and 20 mm vertical.

**Luminance** Standard desktop displays designed for use inside buildings with artificial lighting exhibit maximum luminance around 250 cd/m<sup>2</sup>, while mobile phones which are meant for outdoor use generally have a maximum luminance between 450 cd/m<sup>2</sup> to 500 cd/m<sup>2</sup>. For each of the reported focal depths, we measured the luminance of our display prototype using a Photo Research PR-715 SpectraScan spectroradiometer with a MS-55 lens attachment. We set the aperture to 0.5° and using a 1 s exposure obtained several readings. Mean values are reported in Table 4.1. A decay in the measured values as the focal distance increases is expected because as our mem-

Table 4.3: Optical configuration parameters for the two prototype implementations as related to Figure 3.2.

Parameter	First Prototype	Second Prototype
Eye Relief	65 mm	40 mm
Aperture	65.5 mm × 80.7 mm	57.25 mm × 50 mm
Display Distance	45 mm	28.629 mm
Membrane Tilt	40°	21°
Screen Tilt	10°	0°

brane stretches, the distance between silver particles increases, which causes a reduced amount of reflected light.

### 4.3 Discussion

Both of the display prototypes employing deformable beamsplitters consist of a set of subsystems that interact to control the image displayed to the user. A comparison of the implemented subsystems in each prototype is shown in Table 4.2. The optical systems for each display were designed using different parameters for the optical design presented in Figure 3.2 with the exact parameters of each prototype aggregated in Table 4.3. Prototype 1 suffered from limitations in optical quality due to an astigmatic aberration requiring an additional lens for correction and limitations in mobility due to the large formfactor and being tethered to a vacuum pump. Prototype 2 improved on the first version by eliminating the need for additional optics, decreasing optical aberrations, increasing field of view, introducing computational blur, and improving mobility by reducing formfactor and removing vacuum pump tethering. However the second prototype does not provide a user focus detector due to limitations in the accuracy of commercially available gaze trackers.

## CHAPTER 5: USER FOCUS DETECTION

As explained in Chapter 2, to be useful in a wide variety of situations, a varifocal display requires some form of user focus detection. Varifocal displays address VAC by having only one dynamic focal depth image plane which may be set to follow the user’s focus meaning they have the requirement of continuous measurement of the focal state of the user.

### 5.1 Measuring Focal State

There are two methods for measuring the focal state of a user: the direct method and the indirect method. The direct method works by measuring the light which has passed through the crystalline lens of the eye and bounces off the *fundus*, the surface furthest from the pupil. Autorefractors image a known illumination pattern in multiple axes and optical powers to determine focal state (Cornsweet and Crane, 1970). Shack–Hartmann wavefront sensors measure the wavefront of light after it has bounced off the fundus to measure the focus of the eye (Liang et al., 1994). Both methods are generally performed using infra-red light.

While research prototype NEDs sometimes use the direct method in the lab, there are several barriers for implementation in commercial displays. First, both direct detection methods require on-axis imaging of the eye, which necessitates more complex optical setups than seen in traditional NEDs severely limiting their utility for AR applications. Additionally, many recent NEDs already contain gaze tracking systems, which enables the indirect method for no additional hardware cost. For these reasons, direct measurement of the focal state of the eye is typically foregone in favor of the indirect method.

The indirect method leverages the HVS’s coupling of vergence and accommodation (Fincham and Walton, 1957; Martens and Ogle, 1959), and determines focal state by measuring the ver-

gence distance of the eyes. This gaze-based method leverages more commonly available gaze tracking hardware and, by tracking both eyes, can compute the 3D fixation point, from which the focal depth may also be known. Gaze tracking hardware has several benefits including much higher sampling rates, commonly above 60 Hz, but can be as high as 1000 Hz, and off-axis tracking capabilities. However, because it does not measure the focal state directly and is based on two parallel systems, any error in determining the vergence has a compounded effect on the error in determining the focal state. It is one of the primary goals of this chapter to describe and characterize the effects of gaze-tracking accuracy on focal state accuracy.

It must be noted that if there is dense depth information for the user's visible environment, an alternative indirect method based on scene depth may be employed. By intersecting the gaze direction with scene geometry, a fast and accurate focal distance may be determined. However this method has a severe limitation in regions near depth discontinuities, in that when the error in gaze tracking accuracy overlaps a discontinuity, the incorrect depth may be displayed. Thus a hybrid methodology will provide the best results; relying on scene depth for good candidate depths, and verifying the correct depth by calculating the 3D fixation point.

## 5.2 Determining Error

### 5.2.1 Calculating Distance of Fixation

Most gaze trackers can provide  $\theta$  azimuth and  $\phi$  altitude data, meaning the 3D gaze vector takes the form

$$G_3 = (x, y, z, \theta, \phi) \quad (5.1)$$

where  $x$ ,  $y$ , and  $z$  describe the location of the center of the eye. Note that most gaze trackers do not report eye location, so external calibration must be performed, such as simply measuring IPD, which is typically done for display calibration. Calculating the fixation point is simply finding the intersection of two gaze vectors. Two vectors in 3D space rarely intersect, and while there are several methods of calculating the median of the line of closest approach, the problem

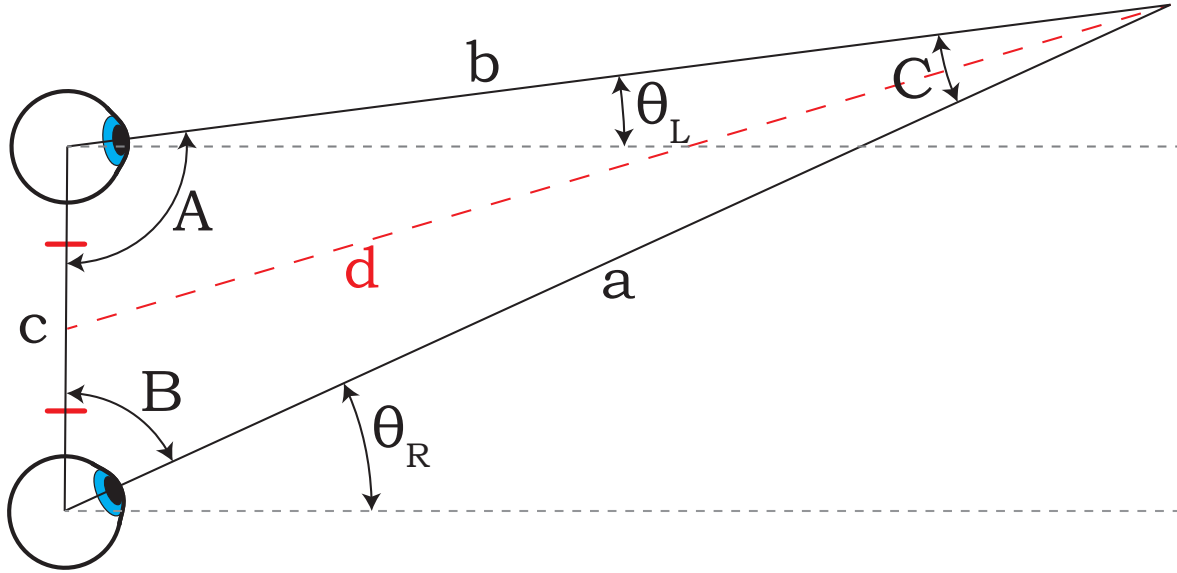


Figure 5.1: Calculating the distance of fixation ( $d$ ) using trigonometric methods.

at hand is much simpler. 3D vectors projected onto a plane — assuming they aren't parallel or diverging — will have an intersection point which can easily be found. The plane intersecting the centers of the eyes and following the head rotation presents a logical choice. With head mounted eye-trackers, head rotation is already included in the gaze vector and assuming the correct gaze-tracker coordinate space, the projection is as simple as dropping the vertical position and altitude from the vector and defining our coordinates such that the eye centers lie along the  $x$  axis, making

$$G = (x, z, \theta) \quad , \quad z = 0 \quad (5.2)$$

the 2D gaze vector equation. Thus it is only horizontal gaze tracker error and IPD error that affect the accuracy of depth calculation.

Given that eye tracking error is generally reported in degrees and the form of the vector, trigonometry provides the most straight forward means of converting tracking error to depth error. With the two gaze vectors  $G_L$  and  $G_R$  for left and right eyes respectively, a triangle is generated where the eye center locations form two corners, with angles  $A$  and  $B$  where

$$A = 90 + \theta_L \quad (5.3)$$

$$B = 90 - \theta_R \quad (5.4)$$

meaning that

$$C = 180 - A - B = \theta_R - \theta_L \quad (5.5)$$

with  $C$  being the angle for the triangle at the point of fixation. Having  $c$  which is the IPD, we can use the Law of Sines to solve the other sides

$$\begin{aligned} a &= \frac{c}{\sin C} * \sin A \\ &= \frac{c}{\sin (\theta_R - \theta_L)} * \cos \theta_L \\ &= c * \cos \theta_L * \csc (\theta_R - \theta_L) \end{aligned} \quad (5.6)$$

$$\begin{aligned} b &= \frac{c}{\sin C} * \sin B \\ &= \frac{c}{\sin (\theta_R - \theta_L)} * \cos \theta_R \\ &= c * \cos \theta_R * \csc (\theta_R - \theta_L) \end{aligned} \quad (5.7)$$

and from Lawes (2013) the median of  $c$ , or cyclopiian distance,  $d$  can be calculated

$$\begin{aligned} d &= \frac{1}{2} \sqrt{2(b^2 + a^2) - c^2} \\ &= \frac{1}{2} \sqrt{2(c^2(\cos^2(\theta_R) + \cos^2(\theta_L))\csc^2(\theta_R - \theta_L)) - c^2} \\ &= \frac{1}{2} \sqrt{c^2((\cos(2\theta_R) + \cos(2\theta_L) + 2)\csc^2(\theta_R - \theta_L) - 1)} \end{aligned} \quad (5.8)$$

giving us the distance to the fixation point from which accommodation is determined by converting to dioptral distance which gives the focal power  $p$ .

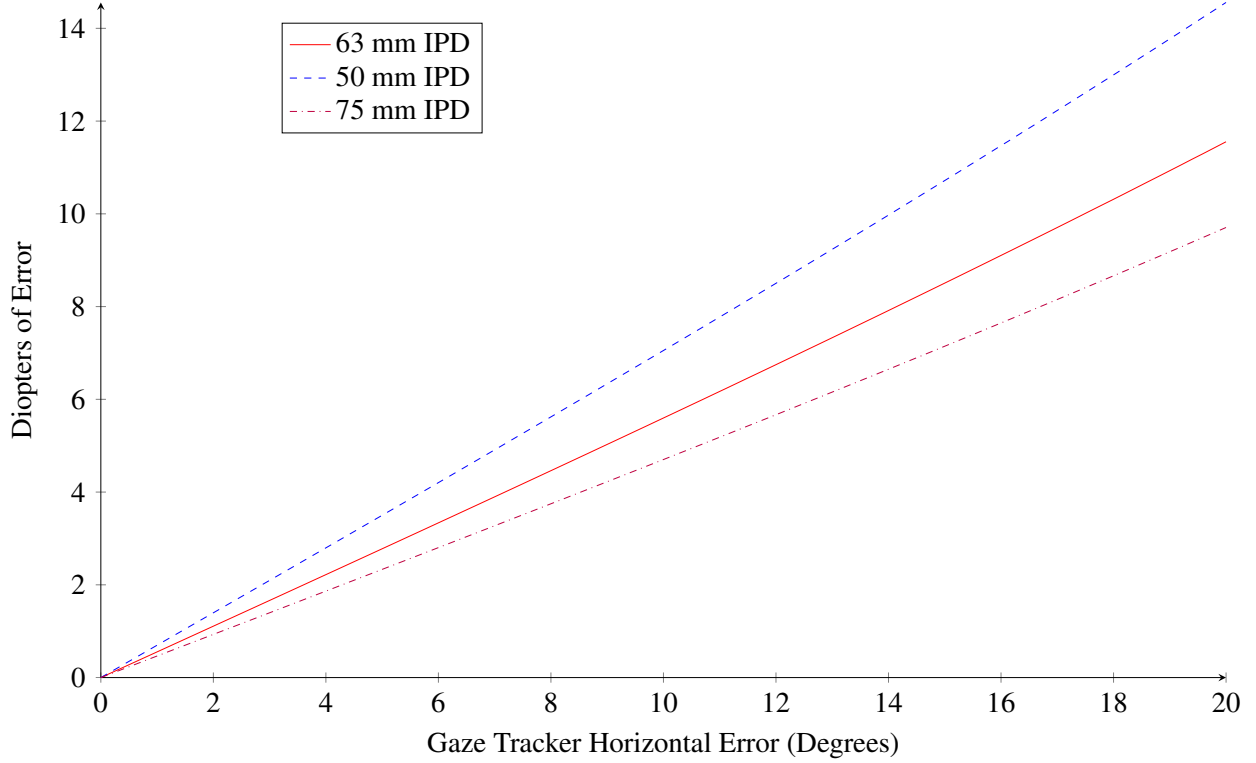


Figure 5.2: Worst case focus error for a gaze tracker with horizontal error for the average American adult IPD (63 mm) and coverage for the vast majority of adults' IPDs (50 and 75 mm) Dodgson (2004) in the central field of view.

$$p = 1/d \quad (5.9)$$

Using these equations, for a given IPD, or  $c$ , we can plot the uncertainty in  $p$  for all errors in  $\theta_R + \theta_L$ . A graph for 3 different  $c$  values displays the error conversion for the central FOV as seen in Figure 5.2. Slight variations are seen at different eccentricities, which I leave to be characterized in future work.

### 5.2.2 Error Assumptions

Now that we have equations for determining the accommodation, I will examine the different varifocal use cases to determine a specification for each. In order to do so, I must make the assumptions stated here:

- Correct measurement of the user's IPD has occurred.
- The display is properly calibrated for the user's IPD and is presenting correct stereoscopic vergence cues for the depth of the stimulus.
- There is proper intra-eye gaze tracker calibration such that the gaze angles from the separate eye trackers are in the same space.
- Since a stereo display should avoid preventing stereo fusion and inability to focus at all costs, I take a tight bound for ZCSBV by using the minimal  $\pm 1.5$  D.
- As the bounds for ZOC are related to length of exposure, a looser standard can be taken here, which I have done with a value of  $\pm 0.8$  D.
- While there is some variance in the DOF of a user, I must make a decision based on the wide distribution of reported values (Wang and Ciuffreda, 2006). I have selected  $\pm 0.3$  D.

### 5.2.3 Differences Between VR and AR

Two common cases for varifocal near-eye displays, VR and OST AR, have different requirements when it comes to accurately setting the focal distance of the display. Just as in the real world, I claim that in VR, as long as the vergence and accommodation stay within the ZCSBV, the image may be fused and focused. Therefore the display must never instantaneously exceed  $\pm 1.5$  D of error; however long-term use may lead to discomfort and fatigue. If kept within the ZOC, a user in VR will not experience negative side effects due to VAC, so the long-term error

Table 5.1: Required Gaze Tracking Accuracy in Central Field

	Dioptral Range	User IPD		
		50 mm	65 mm	75 mm
ZCSBV	$\pm 1.5$ D	$2.15^\circ$	$2.7^\circ$	$3.22^\circ$
ZOC	$\pm 0.8$ D	$1.15^\circ$	$1.44^\circ$	$1.72^\circ$
DOF	$\pm 0.3$ D	$0.43^\circ$	$0.54^\circ$	$0.64^\circ$



should remain below  $\pm 0.8$  D. Based on the above equations, this means that in the central FOV, the average user requires instantaneous gaze tracking error to remain less than  $2.7^\circ$ , and long-term error to remain less than  $1.44^\circ$  if the display relies wholly on the fixation distance method.

While the ZOC also applies to OST AR, due to the real world being visible, the user has an additional reference point, and can directly compare the real and virtual images. When a virtual object is co-located with a real-world object — one of the chief advantages of AR — matching the focus of the virtual to the real can greatly improve perceived image quality because virtual content will not appear out of focus as the user fixates on the real world. This means that at most, the virtual object focus must be within the DOF of the user; which is assumed to be  $\pm 0.3$  D for ease of calculation. With this requirement, gaze tracking must maintain an error of less than  $0.54^\circ$ .

### **5.3 Evaluation of Near-Eye Gaze Trackers**

Now that we know how much accuracy is required for different types of varifocal displays, I will review gaze tracking techniques to evaluate their fitness for the task. Many commercial solutions exist, however it is beyond the scope of this work to evaluate them. The Eye Movement Equipment Database, created by Dr. David Wooding of the Applied Vision Research Unit of the University of Derby is now maintained by the Applied Vision Research Centre of Loughborough University and provides links to many manufacturers of eye tracking hardware<sup>1</sup>.

#### **5.3.1 Intrusive Eye Gaze Trackers**

An intrusive eye gaze tracker is any tracker which requires physical contact with the subject while measuring the gaze direction. There are two methods in practice, scleral search coils (SSC) and electrooculography (EOG).

---

<sup>1</sup> [https://www.lboro.ac.uk/microsites/research/applied-vision/projects/vision\\_resources/emed.htm](https://www.lboro.ac.uk/microsites/research/applied-vision/projects/vision_resources/emed.htm)

SSC is the gold standard of eye tracking techniques. A coiled wire embedded in a contact lens is suctioned to the scleral region of the eye and the voltage induced in the coil by the surrounding electro-magnetic field is measured (Robinson, 1963; Collewyn et al., 1975). They provide extremely high accuracy of  $0.08^\circ$ , and meet all varifocal requirements, however their dependence for custom and extremely intrusive hardware limits their applicability for widespread deployment.

EOG operates by placing electrodes in the eye region and measuring skin potentials. By detecting differences in the skin potentials, eye motions can be detected and filtered from other signals. It is very common in clinical applications due to its lower cost and less intrusive nature. Accuracies around  $2^\circ$  are reported, so the method is insufficiently accurate for varifocal displays (Kaufman et al., 1993).

### **5.3.2 Non-intrusive Eye Gaze Trackers**

Several methods using lights and camera sensors provide methods of gaze tracking that do not require physical contact with the subject.

Infrared oculography (IROG) employs IR LEDs and phototransistors arrayed near the eye. By illuminating the eye using IR LEDs, the phototransistors are able to detect the differences in diffuse reflections between the sclera, iris, and pupil. This results in a voltage difference that is proportional to the angular deviation of the eye (Reulen et al., 1988). The reported accuracy of  $0^\circ 2'$  make it a good candidate for varifocal display gaze tracking, and its hardware could be adapted to work with head-mounted displays.

The dual Purkinjie method developed by Cornsweet and Crane use the first and fourth Purkinjie reflections to separate translational head motion from rotation eye motion. Purkinjie reflections are specular reflections from the different surfaces of the eye (Cornsweet and Crane, 1973). The results are an impressive  $0^\circ 1'$  accuracy, however the complex optical layout and hardware requirements lead to larger formfactors which limit its applicability to varifocal displays.

Table 5.2: Reported Gaze Tracking Accuracy

	Accuracy (°)
Scleral search coils (SSC)	0.08
Electrooculography (EOG)	2.0
Infrared oculography (IROG)	0.033
Dual Purkinjie	0.0166
Video oculography (VOG)	1.0
Machine learning (ML)	2.06

There is a wide range of video oculography (VOG) techniques which use image features in an attempt to locate the center of the eye in an image, and use calibration mapping to convert the pupil location to a gaze direction. With names such as Starburst, SET, ExCuSe, or ElSe, they employ distinct algorithms capable of locating the pupil center in an image (Li et al., 2005; Javadi et al., 2015; Fuhl et al., 2015, 2016). They have reported accuracy of as good as  $1^\circ$ . Having simple hardware, this approach could make a good candidate for varifocal displays if the algorithms can achieve better accuracy.

The recent resurgence of machine learning (ML) research has also affected gaze tracking techniques. Baluja and Pomerleau (1994) and Tew (1997) introduced neural networks trained on a combination of near-eye images and synthetic images. Current state of the art results report accuracies in the range of  $4.5^\circ$  (Park et al., 2018a,b) and  $2.06^\circ$  (Kim et al., 2019). If the accuracy continues to improve, in the near future ML techniques may have the accuracy required for varifocal display gaze tracking.

### 5.3.3 Remote Eye Gaze Trackers

Several techniques and applications exist for remote eye gaze trackers (REGT). The obstruction of the eyes due to wearing the head-mounted display prevents REGT from functioning. Thus I will not review them here.

## 5.4 Discussion

Given the perception-based requirements of PFA, DOF, ZCSBV, and ZOC, I calculated the accuracy needed to present varifocal stimuli in different conditions maintaining a natural viewing state. The required accuracy varied based on the IPD of the subject, but for a median user was  $1.44^\circ$  for VR and  $0.54^\circ$  for AR OST.

Unfortunately, many of the current eye tracking technologies found in literature do not have the required gaze accuracy for driving varifocal NEDs. Of those that do have sufficient accuracy, some require intrusive hardware that make them implausible, while others have complex optical systems which are cost-prohibitive. Of the existing techniques, only IROG meets both the accuracy and hardware requirements for driving varifocal near-eye displays.

I hope that future work in eye tracking may combine an accurate and convenient eye tracking method, such as IROG, with a varifocal NED to provide an improved, more useful varifocal NED.

## **CHAPTER 6: PERCEPTUAL VALIDATION AND ENHANCEMENT**

In this chapter, I present a series of perceptual user studies designed to validate the varifocal capabilities of the prototype displays and explore possibilities of employing varifocal displays in enhancing human perception. As our display is capable of matching the focus of virtual content to real-world objects, we designed and carried out an experiment for determining the benefits to task performance when the focus matches as compared to a non-matching scenario which is described in Section 6.1. Of particular interest is the amount of time taken by the focus changing mechanism inside the human eye as it changes fixation from one depth to another as described in Section 2.2. We also designed and performed a series of experiments in seeking to speed up this response by over-driving the visual input as described in Sections 6.2 through 6.4.

### **6.1 Monocular Acuity Study**

The goal of our first experiment was to verify whether the accommodation support of our first prototype works well, and if users can benefit from it while performing visual acuity task in a monocular viewing scenario.

#### **6.1.1 Hypothesis**

A correct accommodation will allow users to resolve higher spatial details and perform high-acuity mixed-reality tasks with better accuracy.

#### **6.1.2 Experiment Configuration**

The physical configuration consisted of a chin and forehead rest on which was mounted our first prototype display near the participants' eyes, a controller for inputting responses, and

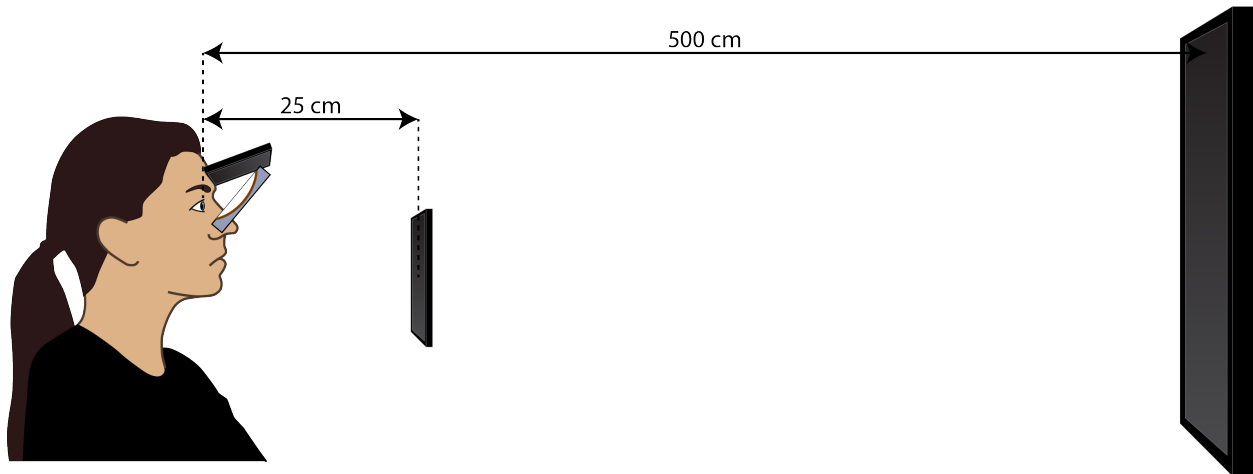


Figure 6.1: Hardware configuration for experiment described in Section 6.1. Two distant displays at depths of 25 cm (4 D) and 500 cm (0.2 D) and our first display prototype are aligned before the subject.

Physical Far						
NED/Virtual						
Physical Near						
Physical	Far	Far	Far	Near	Near	Near
Camera Focus	Far	Far	Near	Near	Near	Far
NED/Virtual	Far	Near	Near	Near	Far	Far

Figure 6.2: Series of photographs showing example stimulus as seen by a participant during our experiment. Labels below each photograph indicates focal state of our camera, physical location of the display, and the depth of the virtual image.

two LCDs located at 0.25 m (Adafruit Qualia, 9.7", 2048 px × 1536 px, 23.5 cpd, 60 Hz) and 5 m (Sharp Aquos Quattron LC-70LE732U, 70", 1920 px × 1080 px, 54.3 cpd, 60 Hz) from the viewer. An illustration of the hardware layout can be seen in Figure 6.1.

### 6.1.3 Stimuli

Each stimulus was comprised of a pair of white Landolt C shapes shown on a black background. Example stimuli can be seen in Figure 6.2. The location of the gaps was either on the top or the bottom corresponding to the up and the down orientation of the shape. The shapes were

separated by 2 visual degrees, and each of them spanned  $0^{\circ}30'$  which imposes the gap size of  $0^{\circ}6'$ , where a normal 20/20 eye can identify a gap size of  $0^{\circ}1'$ . Since through our NED calibration, the focus state has been precisely setup for each trial, we opted for the larger gap size so that the user response is immediate and effortless, additionally, it is less affected by lower display contrast, limited spatial resolution, and possibly imperfect luminance adaptation with respect to the requirements of a standard visual acuity test. One Landolt shape was presented on one of the two distant displays, while the other Landolt shape was presented on our NED with a focal distance either matching the distance to the physical screen or a different one to simulate a lack of a correct accommodation cue. The range of considered focal distance offsets was 0.2 D to 5 D. For the screen located at 0.25 m, we moved the virtual object further from the observer, while for the screen located at 5 m, we moved the virtual image closer to the observer.

#### **6.1.4 Participants**

Twelve subjects (2 F, 10 M, 20 to 34 years of age) that had a normal or corrected-to-normal vision, took part in the experiment. To keep the eye of the subject inside the eye box of our NED, all participants used a chin and forehead rest.

#### **6.1.5 Procedure**

For each trial, the subject was asked to monocularly fixate with the right eye on one of the physical screens. To assist with this, a simple math equation was displayed on the screen using a font height of  $0^{\circ}15'$ , while nothing was displayed on our NED. The user was asked to push one button if the equation was true and another if it was false. This math task was introduced to control the user fixation and give enough time to correctly accommodate to the distance at which the physical screen was located. Immediately after responding, the stimuli appeared on the reference display and the NED at a location central to the equation. The presentation time of the stimulus was set to 300 ms. The time was chosen such that it was just-enough to perform the visual acuity task, which was determined during a pilot experiment. Note that the time is also shorter than the

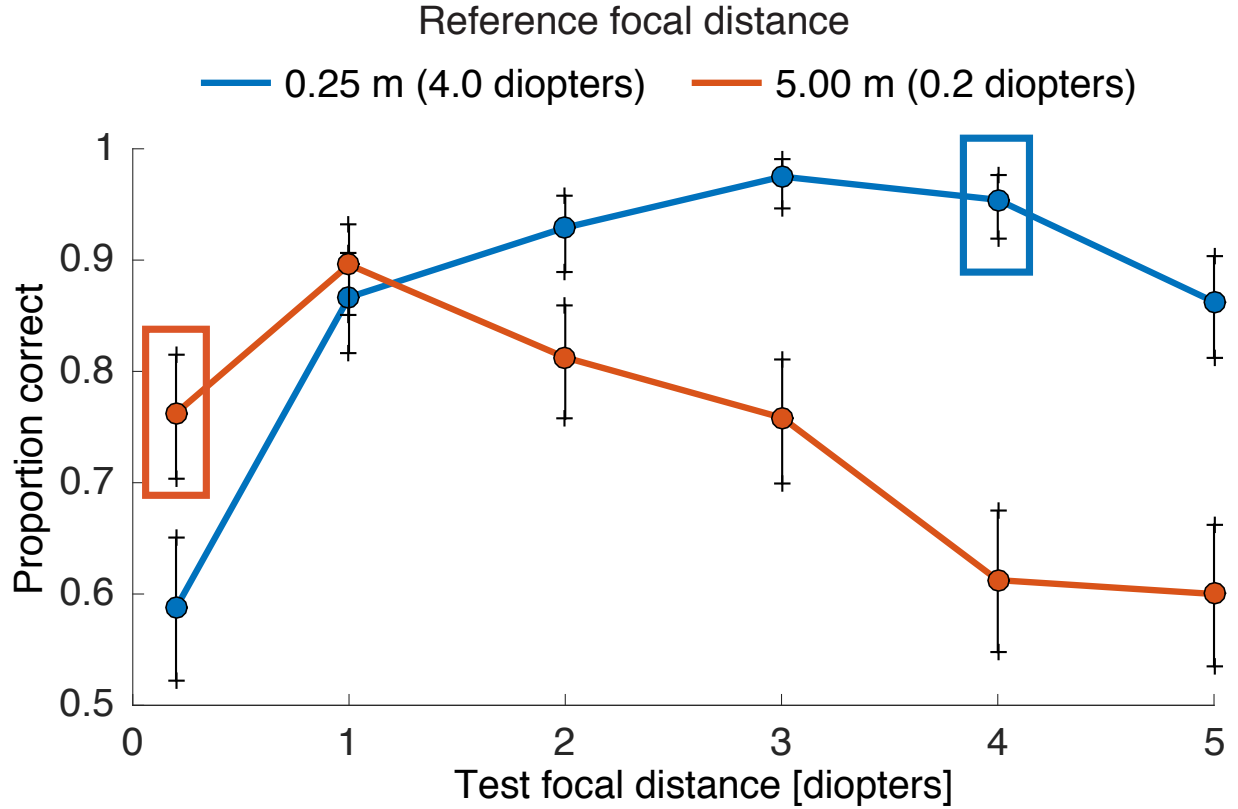


Figure 6.3: The proportion correct as a function of test focal distance of the NED. Two points marked by rectangles are points where the reference and the test distances matched. For such conditions, the performance is expected to be the best. The error bars denote Clopper-Pearson binomial confidence intervals.

latency before the actual change in the eye lens shape is triggered, which I discussed in more detail in Section 2.2. Next, the subject was presented with a blank screen and asked to press a button selecting whether the two patterns just extinguished were of equal or different orientation. Afterwards, the study continued with the next trial. In total, two physical displays and six focus distances for the NED were used in random order, which after 20 repetitions gave the total of 240 trials per participant. Each participant took, on average, 30 min to complete the task.

### 6.1.6 Results

The graph in Figure 6.3 shows the relation of the NED focal distance and the proportion of correct responses for each of the reference displays. We performed a  $\chi^2$ -test to analyze differences between different conditions and found a significance influence of the test focal distance on



the proportion correct for both 0.2 D ( $\chi^2 = 82.7, df = 5, p < 0.001$ ) and 4.0 D ( $\chi^2 = 204.7, df = 5, p < 0.001$ ) references. A post-hoc analysis with Bonferoni correction and significance level equal to 0.05 revealed that the differences between test pairs were significant for all but the following: 0.2-2.0 D, 0.2-3.0 D, 2.0-3.0 D, 4.0-5.0 D for 0.2 D reference and 1.0-2.0 D, 1.0-5.0 D, 2.0-4.0 D, 3.0-4.0 D for 4.0 D reference.

In general, as the test focal distance approached the reference depth, i.e., both stimuli were presented at the the same focal distance, the participants were able to correctly perform the task more often maxing out at 97.5 % and 89.6 % for stimuli at 0.25 m (4 D) and 5.0 m (0.2 D), respectively. The best performance should be located at the points corresponding to the cases where the test and the references focal distances match (see rectangles in Figure 6.3). This can be observed well for the closer physical display. For the further screen, the drop of the performance for the isofocal condition can be explained by a degradation of the image quality due to the strong membrane distortion compensation required for such an extreme focus depth in our first prototype. This made the comparison of relatively small shapes difficult. Except for this particular case, the trend in our measured data follows the expectation, i.e., the participant performance drops with increasing optical focus difference between both displays.

For the reference display at 0.25 m distance (4 D, blue) and our NED set up to focus at 1.00 m (1 D), participants had to compare two shapes at focal distances differing by 3.0 D and had a mean performance of 86.7 %. As our analysis shows, this is a significant drop from the optimal performance when the focus for NED matches the physical screen. Similar observations can be made for the reference display at the distance of 5.00 m (0.2 D, red), where the performance significantly drops to 75.8 % for the NED focused at 0.33 m (3 D) when compared to the case of focusing it at 1.0 m (1 D). Situations like these also occur while using current AR displays with fixed focus distance. From these results, we conclude that the varifocal properties of our device allow improving the user performance in tasks that require simultaneous observation of the virtual and the real worlds.

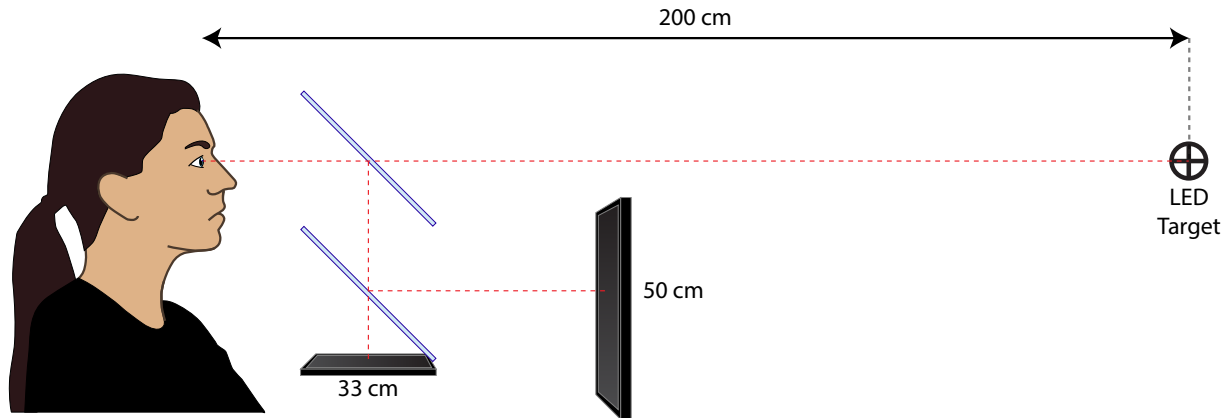


Figure 6.4: Display hardware configuration for running monocular focus over-driving experiment described in Section 6.2.

## 6.2 Over-driving Focus

It takes some time for the human eye to refocus on content at a different depth. For accommodation, the maximum velocity of the change is determined by the amount of change required, so larger changes have larger maximum velocities (Schor and Bharadwaj, 2005). This may not always be true for disaccommodation (Bharadwaj and Schor, 2006). If we initiate a depth change by presenting an initial stimulus with a depth difference larger than the target depth change, then during the accommodation period back off to the correct final depth, we may see a speed-up in the perception of the stimuli. This over-driving of the focal system could lead to enhanced perceptual capabilities of our display when compared to the real world, and perhaps improved performance of other varifocal AR displays when compared to the real world. In this and the following experiments, we attempt to induce this over-driven behavior in subjects of our experiment by presenting visual stimuli and measure results based on performance in a time-constrained identification task.

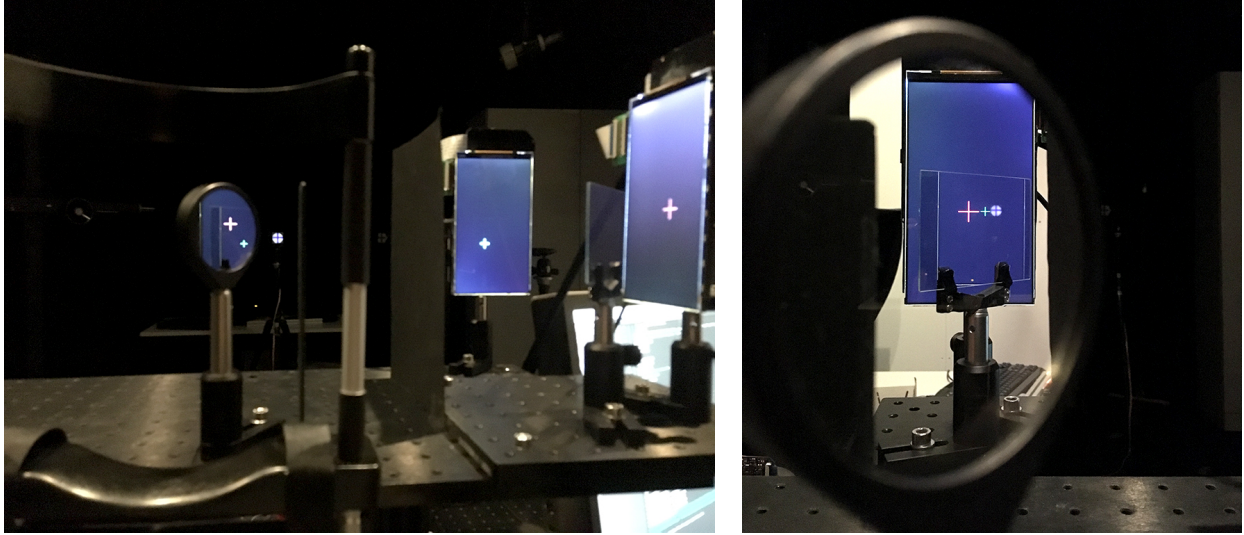


Figure 6.5: View of physical display hardware for running monocular over-driving focus experiment described in Section 6.2. Left) beamsplitters, LED target, and LCDs are visible. Right) slightly offset view through the display with calibration markers visible.

### 6.2.1 Hypothesis

A user given over-driven focal cues in certain circumstances can perceive virtual content faster than if correct focal cues are provided.

### 6.2.2 Experimental Configuration

The physical configuration consisted of a chin and head rest, a controller for input, an LED target positioned 200 cm from the user, and two LCDs positioned 50 cm and 33 cm from the user visible via planar beamsplitters as seen in Figure 6.4 and Figure 6.5. The first beamsplitter was positioned in front of the participant's right eye, while the left eye was obstructed by means of a blacked-out barrier.

### 6.2.3 Stimuli

A preparation stimulus was created by back illuminating the cross-hair target with a diffused LED. The main stimulus was comprised of a single Landolt C shape spanning  $0^{\circ}10'$  which sets the gap size to  $0^{\circ}2'$ , where a normal 20/20 eye can identify a gap size of  $0^{\circ}1'$ . With equal proba-

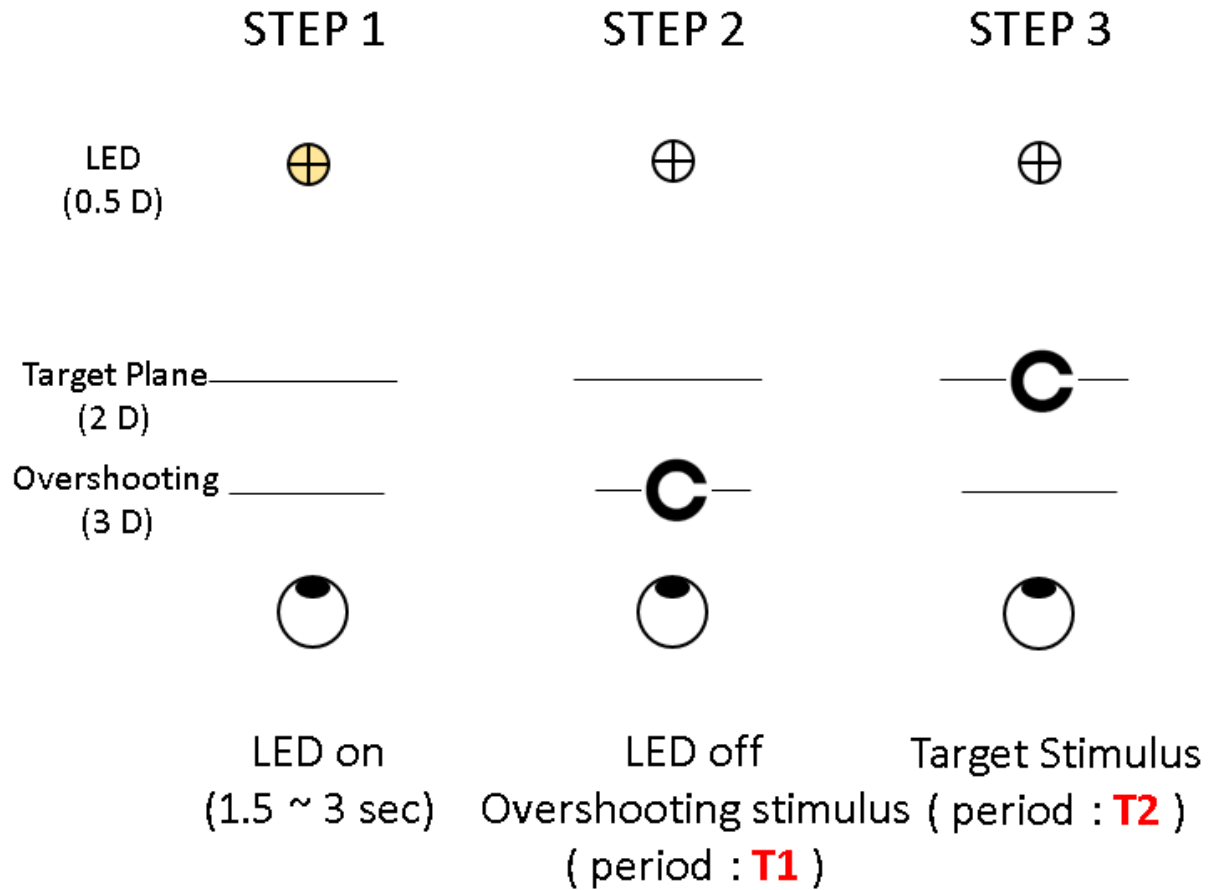


Figure 6.6: Experimental procedure for running monocular over-driving focus experiment described in Section 6.2.

bility, the Landolt C shape was randomly oriented with the gap in one of the four directions, up, right, down, and left.

#### 6.2.4 Participants

Four subjects (all M, 25 to 35 years of age) that had a normal or corrected-to-normal vision, took part in the experiment. To keep participants inside the eyebox of our multiplane display, all participants used a chin and forehead rest.

### 6.2.5 Procedure

The goal of the experiment was to find the detection threshold for identifying the orientation of the Landolt C for four independent test cases. For each test case, we ran a virulent PEST staircase procedure as described by Taylor and Creelman (1967) and Findlay (1978). All four staircase procedures were interleaved to run in parallel.

A single trial — described pictorially in Figure 6.6 — consisted of 4 phases:

- Preparation phase, this phase was responsible for ensuring correct initial focus at 200 cm (0.5 D) by illuminating the LED target at the initial depth. This stimulus was illuminated for a random period between 1.5 s and 3 s.
- An overdrive phase in which the main randomly oriented Landolt C stimulus was displayed on the 33 cm (3 D) screen for a specified number of frames  $T1$ .
- The main stimulus phase where the same Landolt C stimulus from the overdrive phase was displayed for a variable number of frames  $T2$  at the destination focus depth on the 50 cm (2 D) screen.
- A response phase where as soon as the main phase time expired, the stimulus disappeared. This phase continued until the user responded with the orientation of the Landolt C stimulus.

After the response was given, the next trial began. The variation between the four test cases was the length of time the over-driven stimulus was visible, or  $T1$ . Times corresponding to the four cases are 0 ms, 67 ms, 133 ms, and 200 ms. The 0 ms test case corresponds to a non-over-driven control case where the overdrive phase is skipped entirely. Each PEST staircase procedure, before ending, independently finds the perceptual threshold for the free variable  $T2$ , which was the number of frames the stimulus was visible in the main phase before disappearing at the start of the response phase.

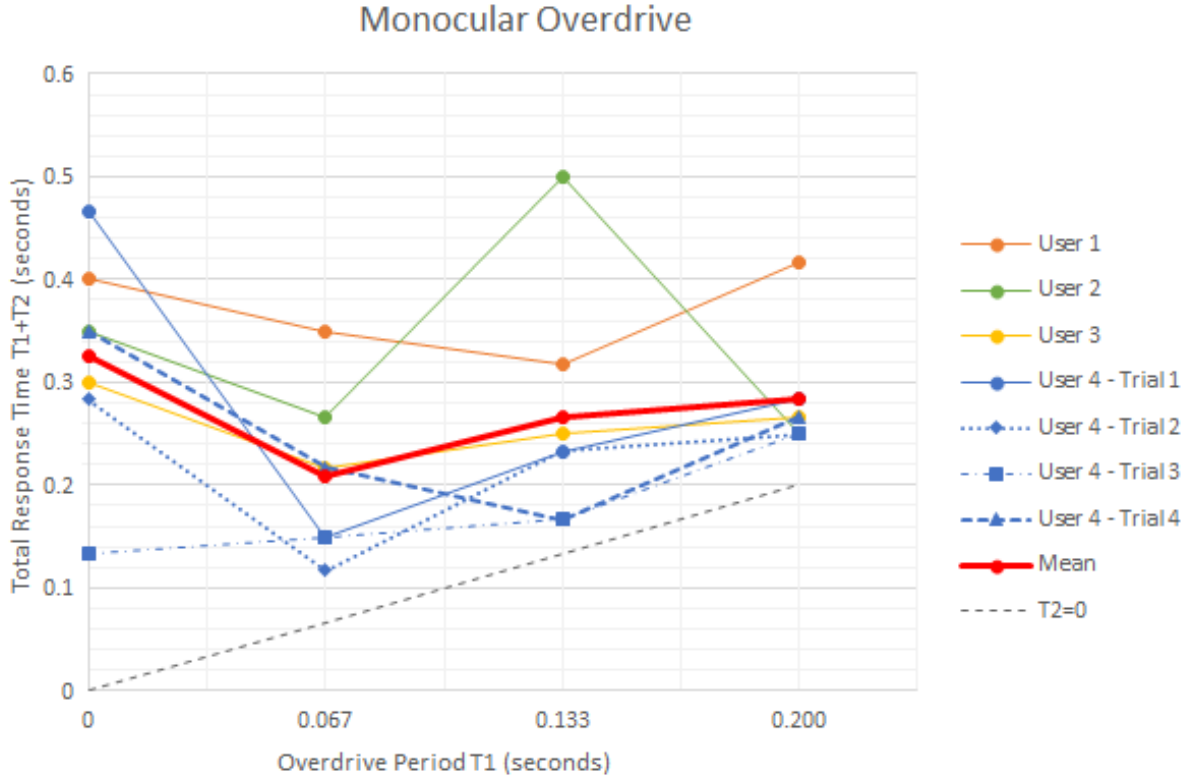


Figure 6.7: The results for the monoscopic over-driving focus experiment described in Section 6.2. Each point represents the threshold detected from an entire staircase procedure. The 4 test cases with differing overdrive periods  $T_1$  are plotted on the horizontal axis, while the total time the stimulus was visible  $T_1 + T_2$  including both the overdrive and main phase times are plotted vertically.

### 6.2.6 Results

A plot of the results can be seen in Figure 6.7. Each point represents the threshold detected from an entire staircase procedure. The 4 test cases with differing overdrive times  $T_1$  are plotted on the horizontal axis, while the total time the stimulus was visible  $T_1 + T_2$  including both the overdrive and main phase times are plotted vertically.

Our hypothesis can be tested by comparing the  $T_1 = 0$  control case on the far left to the other cases with some amount of overdrive time for the same user. As can be seen, for each user, with the exception of user 4 trial 3, there was at least one case where the total time to identify the stimulus was improved with the additional overdrive time. This indicates that not only is

over-driving the accommodation response possible, but that it can speed up perception of visual stimuli.

In the single case that did not exhibit a speed up, user 4 trial 3, we speculate that the control response time was so quick that there wasn't very much room for improvement, as can be seen by how close it is to the  $T_2 = 0$  boundary line. All the other trials from the same user exhibited the expected behavior, so trial 3 is viewed as an outlier.

### **6.2.7 Discussion**

While we didn't run enough users through the experiment to make any generalized claims based on the results, the overall result showed the promise of over-driving focus, which merited further study. In this study we examined the one-eye case which only accounted for speeding up the eye's accommodation. While this result may be useful for users with depleted vision in one eye, such as those with Amblyopia, or those with stereoblindness, we wanted to know if the same speed-up could be exhibited in binocular situations where both accommodation and vergence are operating in concert. So we developed and ran the experiment described in Section 6.3 below.

## **6.3 Over-driving Focus and Vergence**

With the promise of enhancement to perception indicated by the previous accommodation-only study for a single eye, we sought to extend the principle to both eyes when vergence and accommodation work together. Thus we performed the following experiment.

### **6.3.1 Hypothesis**

A user given over-driven focal and vergence cues can perceive virtual content faster than if correct focal and vergence cues are provided.

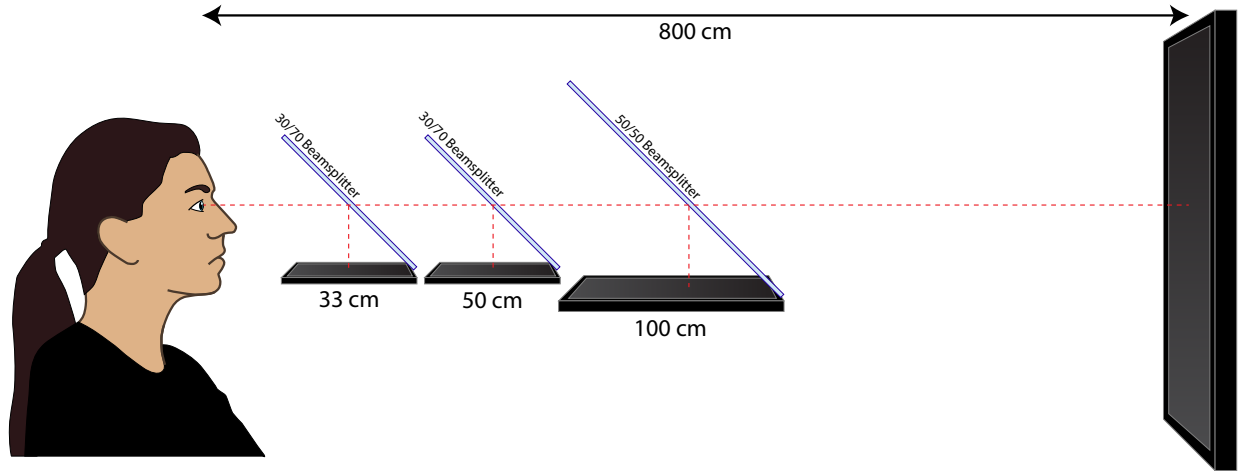


Figure 6.8: Display hardware configuration for running binocular accommodation and vergence over-driving experiment described in Section 6.3. Four displays at depths of 33 cm (3 D), 50 cm (2 D), 100 cm (1 D), and 800 cm (0.125 D) combined with a series of planar beamsplitters with increasing reflectance ratios.

### 6.3.2 Experimental Configuration

The physical configuration consisted of a chin and head rest, a controller for input, and a 4-depth multiplane display which employed a combination of LCDs and planar beamsplitters. The 3 beamsplitters and 4 LCD screens were distributed such that the depth of the images were 33 cm, 50 cm, 100 cm, and 800 cm distant from the user, as can be seen in Figure 6.8 and Figure 6.9. Stereoscopic views were not supported by the display because the images were visible to both eyes, meaning depth fusion doesn't work, so virtual objects could only be presented at one of the 4 fixed depths.

### 6.3.3 Stimuli

For this experiment the preparation stimulus consisted of a series of randomized true/false math equation with the form:

$$x + y = z \quad (6.1)$$

where  $x$  and  $y$  were single digit integers and  $z$  was the answer. When the equation was false,  $z$  was altered by adding  $\delta$  where  $\delta \in \{n \in \mathbb{Z} \mid -2 \leq n \leq 2, n \neq 0\}$ . The equations had an even





Figure 6.9: View of physical display hardware for running binocular vergence and accommodation over-driving experiment described in Section 6.3. Top Right) . Top Left) .

probability of being true or false. The equation was centered along the line of sight with the characters subtending  $0^{\circ}9'$  visual angle each and the entire equation subtending a horizontal visual angle of  $0^{\circ}45'$  or  $0^{\circ}54'$ .

The main stimulus consisted of a vertical column of 12 Landolt C shapes, all oriented with the gap in the same one of the four directions: up, right, down, and left. Each Landolt C spanned  $0^{\circ}10'$  making the gap size  $0^{\circ}2'$ . The entire column of Landolt C shapes subtended a vertical angle of  $3^{\circ}50'$ . The choice of presenting more than one shape was made to avoid adding an indeterminate amount of visual scanning time to each trial while the user attempts to locate the relatively small stimulus. Enlarging the stimulus would have decreased the visual acuity required to correctly identify the orientation which would have been detrimental to the results. A verti-

cal column was chosen to avoid any vergence ambiguity that multiple horizontal stimuli would introduce.

During the response phase, an additional response stimulus was displayed allowing the user to continue to fixate at the target depth while they selected their response. The response stimulus was composed of a vertical column of 12 circular shapes co-located with the main stimulus. Thus the only difference between the main stimulus and the response stimulus was the gap in each Landolt C, which simply disappeared when the response phase was entered.

#### **6.3.4 Participants**

Eight subjects (2 F, 6 M, 23 to 44 years of age) attempted the experiment. 2 subjects were eliminated (2 M) for not having normal or corrected-to-normal vision at both 800 cm and 33 cm meaning six subjects total completed the study. To keep participants inside the eyebox of our NED, all participants used a chin and forehead rest.

#### **6.3.5 Procedure**

We began the procedure by performing a one-time-per-subject display alignment calibration, which aligns the subject's eyes vertically with our display and performs an IPD adjustment. Then, user acuity was tested with a 3-down/1-up staircase procedure for detecting the orientation of a Landolt C shape using the size of the shape as the free variable. Users with acuity better than 20/30 were allowed to continue. At that point a basic training sequence familiarized the subject with the task and procedure and tested the user to ensure correct responses with a series of practice rounds. Three correct responses were required to continue on to the main experiment.

The goal of the main experiment was to find the detection threshold for identifying the orientation of the Landolt C on seven independent test cases. For each test case, we ran a virulent PEST staircase procedure as described by Taylor and Creelman (1967) and Findlay (1978). All seven staircase procedures were interleaved to run in parallel. Four of the test cases were analogous to the previous experiment, all having the same depth change, but with different over-drive

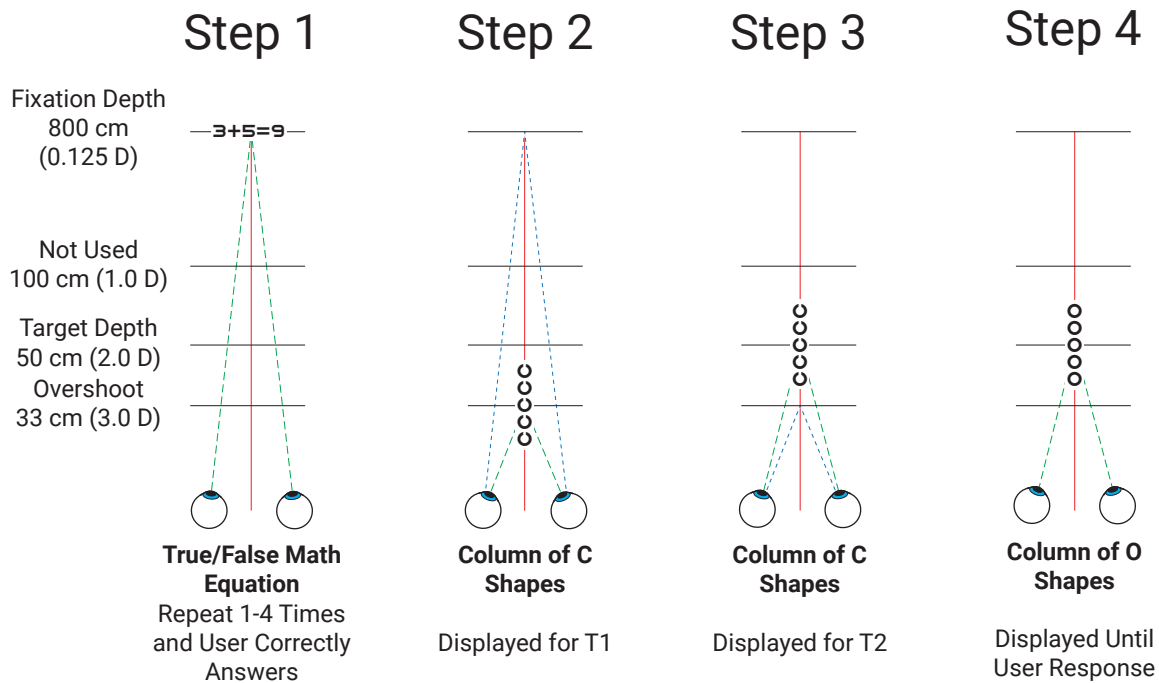


Figure 6.10: Experimental procedure for running binocular over-driving accommodation and vergence experiment described in Section 6.3.

periods. The other three test cases had no over-drive periods, but instead varied the amount of depth change between the preparation stimulus and the main stimulus. By randomizing the amount of depth change, these three extra cases worked to prevent a learning effect where the subject would anticipate the focus change.

A single trial — described pictorially in Figure 6.10 — consisted of 4 phases:

- The preparation phase ensured correct initial focus depth at 800 cm (0.125 D) by presenting a sequence of the preparation stimuli. The initial number of problems in the sequence was determined by a random choice from the distribution (1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4). Each stimulus is shown until the user responds. If a user does not answer the last problem correctly, an additional problem is presented. This randomization was also introduced to prevent a learning effect where the subject would anticipate the focus change.
- Between the preparation phase and main stimulus phase, an overdrive phase would present the main stimulus on the 33 cm (3 D) screen for a specified number of frames  $T1$ .

- The main stimulus phase where the main stimulus was displayed at the target depth  $d$  for a variable number of frames  $T2$ .
- A response phase where as soon as the main phase time expired, the main stimulus would be replaced by a response stimulus at the same location. This phase continued until the user responded with the orientation of the main stimulus gap.

After the response was given, the next trial began. The variation between the four main test cases was the length of time the over-driven stimulus was visible, or  $T1$ . Frames corresponding to the four cases are 0 frames (0 ms), 5 frames (~83 ms), 10 frames (~167 ms), and 15 frames (~250 ms). For all four main test cases, the target depth  $d$  is set to 50 cm (2 D). The 0 ms test case corresponds to a non-over-driven control case where the overdrive phase is skipped entirely.

The additional three test cases also have  $T1$  set to zero, meaning no overdrive, however their target depths  $d$  are set as: 800 cm (0.125 D), 100 cm (1 D), and 33 cm (3 D). Each PEST staircase procedure, before ending, independently finds the perceptual threshold for the free variable  $T2$ , which was the number of frames the stimulus was visible in the main phase before disappearing at the start of the response phase. The entire procedure took about 30 min of time to complete.

### 6.3.6 Results

We can glean two sets of results from these experiments. The first is visualized in Figure 6.11, which is a plot of the three additional test cases and the control case from the main test set, or all test cases where  $T1 = 0$ . Again, each point represents the threshold result of an entire staircase. In this plot, the target depth  $d$  makes the horizontal axis, while the thresholded amount of time required to correctly identify  $T2$  is the vertical axis. Here we expect to see that as the amount of focus change increases between the preparation and main stimuli, that the amount of time also increases, and this is exhibited nicely, in an almost linear manner. These data also show that our measures against subjects anticipating the focus change worked. The 0 D case has the lowest latency for all cases.

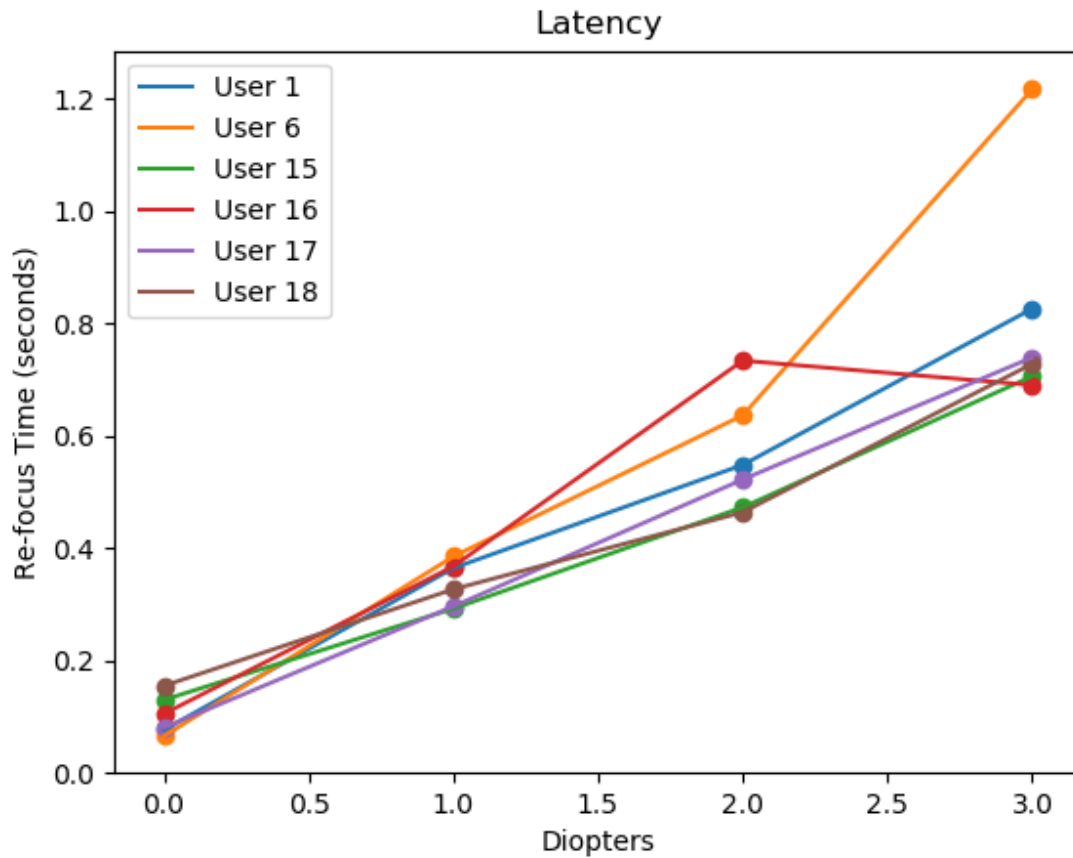


Figure 6.11: The time required to recognize a stimulus after the depth change given in diopters during the over-driving focus and vergence experiment.

The second set of results, similar to the results of the previous experiment, will make or break our hypothesis. By looking at Figure 6.12, we can see no statistically significant result. Unfortunately, our over-drive of vergence and accommodation together did not speed up perception as we thought it would.

### 6.3.7 Discussion

In attempting to over-drive both vergence and accommodation together, our experiment failed. We suspected that one reason for this might be that no change in gaze angle was induced. The testing scenario where fixation was quickly changed from far to near without a saccade is a case that is rarely seen in the real world. Objects are, generally speaking, opaque. So when fixation

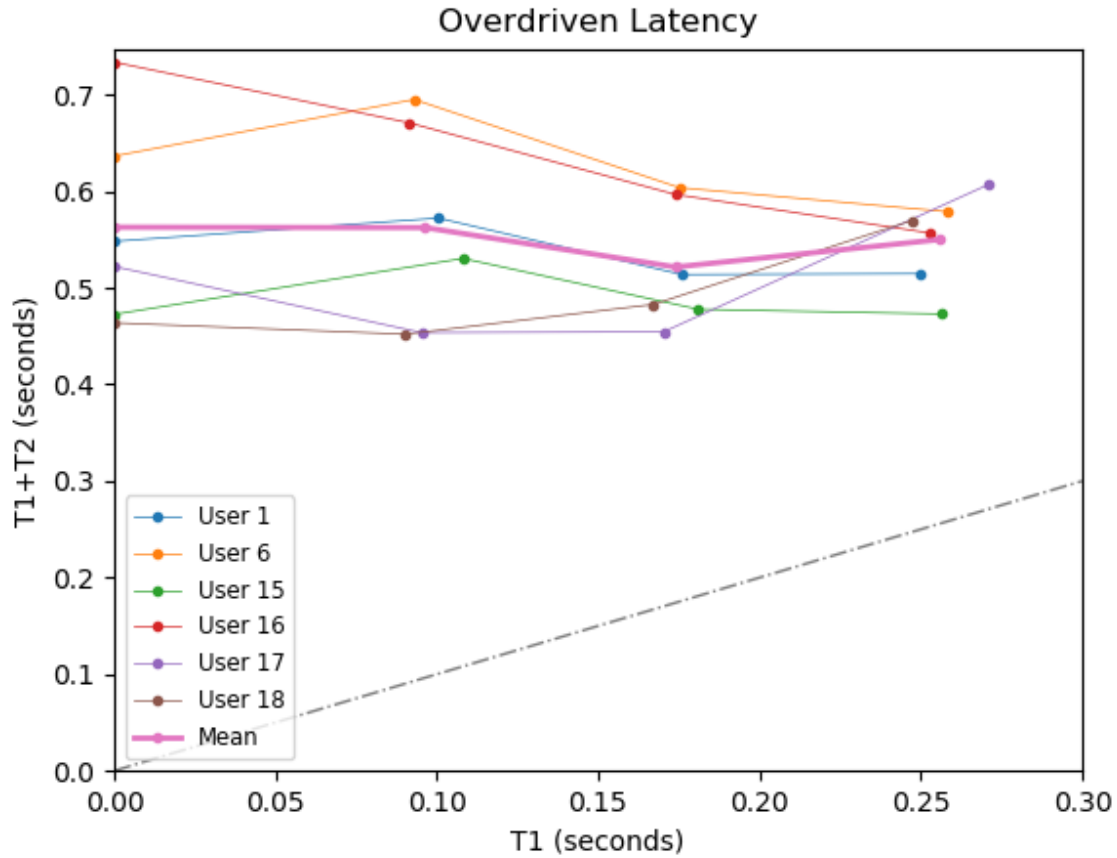


Figure 6.12: The time required to recognize a stimulus after a 2 D depth change when the stimulus is over driven for  $T_1$ .

changes from an object at one depth to an object at a closer depth, it is traditionally accompanied by a change in gaze angle. By testing a rarely seen situation in the real world, we thought we may be selling the idea short, so we devised yet another experiment as described in Section 6.4 below.

#### 6.4 Over-driving Focus and Vergence with Saccade

With the non-results seen in the previous binocular accommodation and vergence study, and still attempting to extend the accommodation-only study for a single eye, a saccade was added during the depth change. Vergence has been shown to be accelerated when accompanied by a saccade, as the eye motions are controlled by different responses and have different speeds.

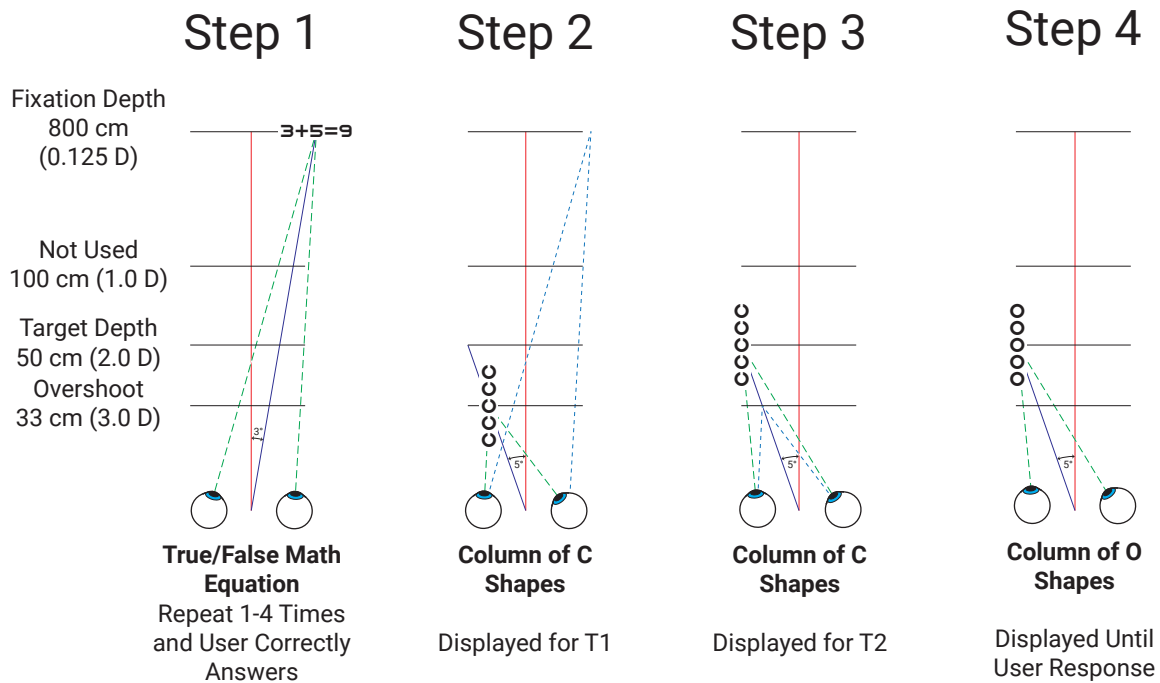


Figure 6.13: Experimental procedure for running binocular over-driving accommodation and vergence with saccade experiment described in Section 6.4.

Thus with this experiment, we investigated if adding a saccade while over-driving the focus and vergence would produce the improved response times.

### 6.4.1 Hypothesis

A user given over-driven focal and vergence cues prior to a saccade can perceive virtual content faster than if correct focal and vergence cues are provided.

### 6.4.2 Experimental Configuration

The configuration of this experiment is exactly the same as described in Section 6.3, which can be seen in Figure 6.8.

### 6.4.3 Stimuli

The same stimuli as in the previous vergence and accommodation study with the following differences:

- Preparation stimulus is located  $3^\circ$  right of the user's central field.
- The main stimulus is located  $5^\circ$  left of the user's central field.
- The response stimulus is similarly located  $5^\circ$  left of the user's central field.

### 6.4.4 Participants

Four subjects (all M, 26 to 35 years of age) took part in the experiment. All other subject attributes are as described for the previous study.

### 6.4.5 Procedure

Exactly the same as previous study detailed in Section 6.3 except with the new stimuli locations as described above, as seen in Figure 6.13.

### 6.4.6 Results

We can analyze the same data as from the last experiment. The latency for test cases without over-drive plot shown in Figure 6.14 shows, even with a few outliers, the similar upward trend: as dioptral change increases, so does the amount of time required for correctly identifying the stimuli. However, with the added saccade, times for the 0 D case were longer, and the results have a larger distribution.

The over-drive results are presented in Figure 6.15. As can be seen there is a stronger effect than in the case with no saccade. With the addition of 10 frames of over-drive, all subjects identified the orientation of the stimuli faster than without the over-drive.



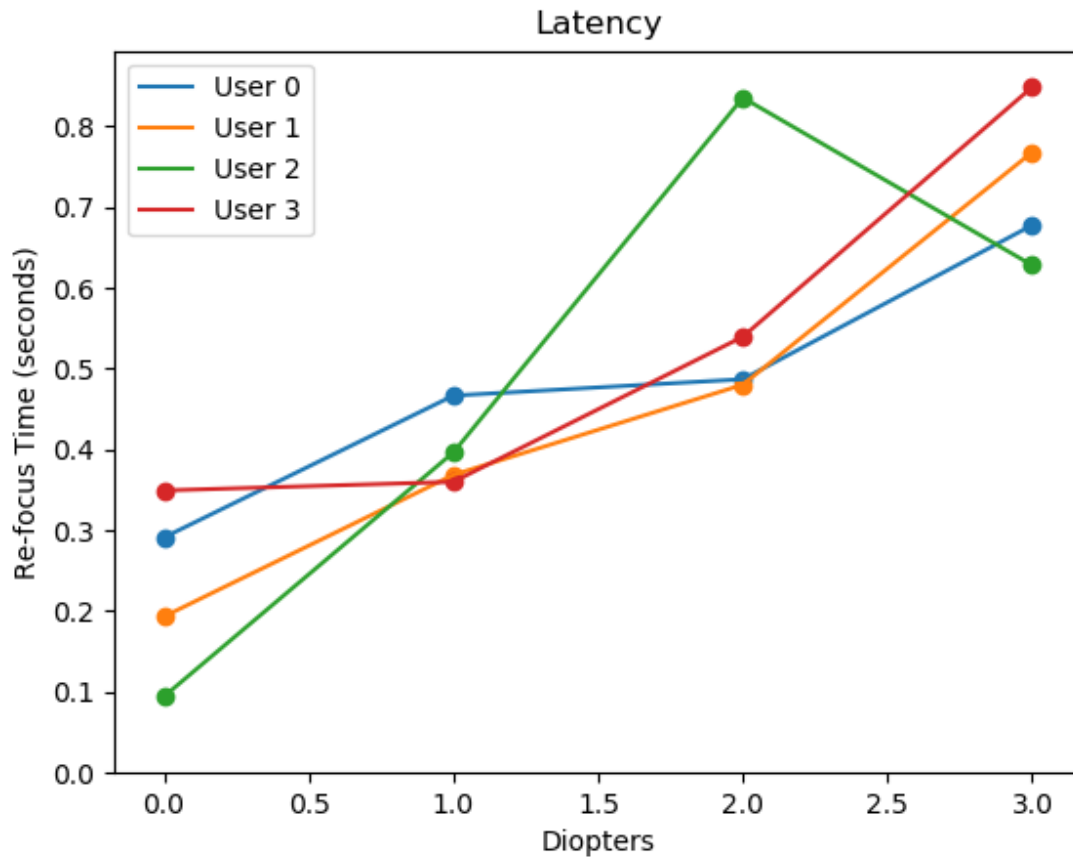


Figure 6.14: The time required to recognize a stimulus after the depth change given in diopters with a saccade of  $8^\circ$  during the over-driving focus and vergence experiment.

#### 6.4.7 Discussion

While this is a good result that indicates our hypothesis is correct, the small sample size and lack of statistical analysis prevents us from fully claiming the result. Further study is required to confirm and further study the possibilities of enhancing human perception by using varifocal or multiplane displays.

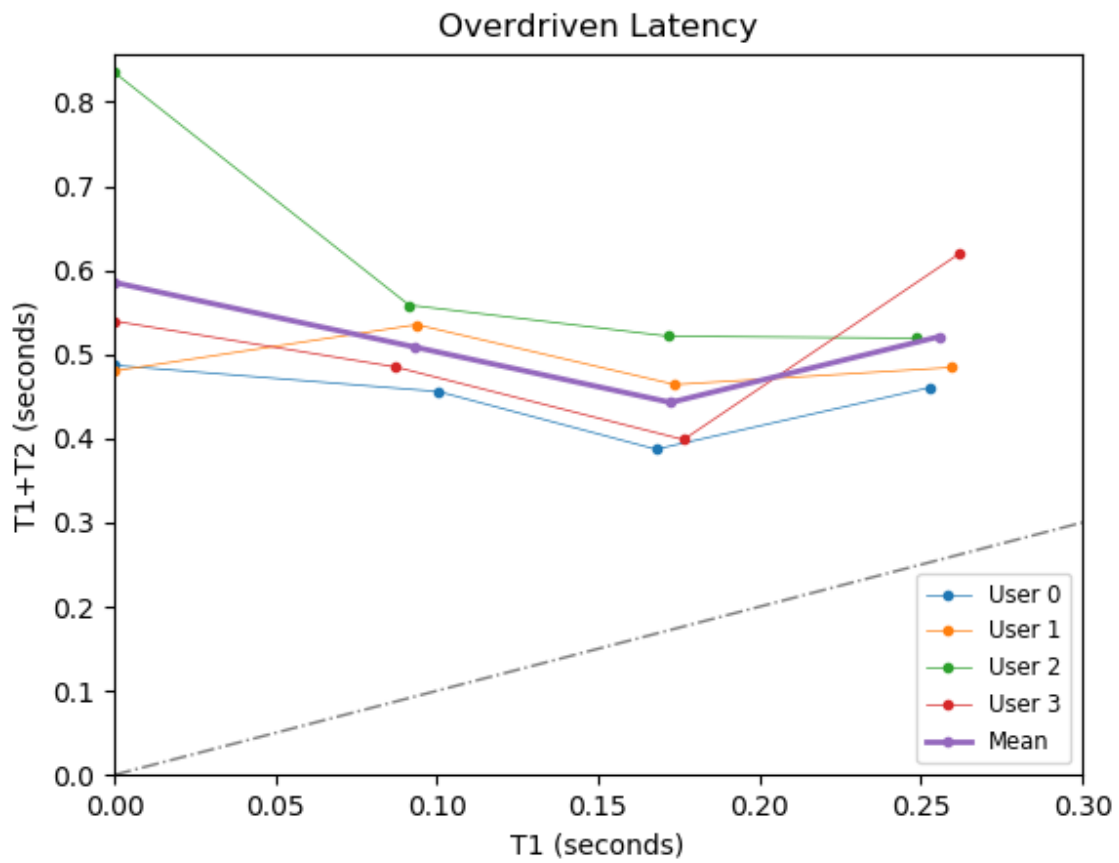


Figure 6.15: The time required to recognize a stimulus after a 2 D depth change and  $8^\circ$  saccade when the stimulus is over driven for  $T1$ .

## CHAPTER 7: SUMMARY AND CONCLUSION

In order to provide a high-quality augmented reality experience, it is crucial to design headsets that are capable of reproducing all visual cues across the entire visual field. In this respect, one of the more challenging tasks is reproducing accommodation cues while providing a wide FOV.

In this work, I have described and characterized a new dynamic optical element, the deformable beamsplitter. This new device is capable of adjusting the focus of the reflected light while not altering the transmitted light; making it ideal for applications in AR NEDs. The key to this solution is a membrane with half-mirror properties which can curvature as needed. AR applications of deformable beamsplitters are evident as the real-world light remains unchanged while light forming the virtual images is controlled.

Two prototype displays capable of providing a wide FOV and large focal range in a good form-factor illustrate these principles, and a user study verified their capabilities. Due to the membrane's deformation characteristics, the focal power can be adjusted using an airtight chamber to provide accommodation cues matching the observer's fixation distance. This approach addresses one source of visual discomfort that is caused by a mismatch between vergence and accommodation. It also improves user task performance as demonstrated in our first experiment. Another unique advantage of our membranes is the fact that they enable a significantly larger field of view when compared to other varifocal designs.

One way of driving the focus for these prototypes is by tracking the gaze direction of the user's eyes and calculating a fixation point. Unfortunately, with the commercially available eye trackers today, the needed combination of accuracy, size, user adaptability, and long-term usability does not exist.

The potential of improving human performance beyond what it is capable of in the natural world is a worthwhile aim. In over-driving the focus of a single eye, users showed a significant improvement over unenhanced abilities. By over-driving both vergence and accommodation during saccadic revergence, the improvements are similarly exhibited for the binocular case. Potential enhancements to human perception enabled by the new optical element seem plausible.

## **7.1 Future Work**

There are several current limitations preventing our deformable beamsplitter displays from being practical consumer devices. The first limitation is the method of actuating the display. Both solutions we enacted on our prototype displays, a vacuum pump and tank with pressure regulator or 8" subwoofers with solenoid valves, are not conducive to mobile consumer-worn devices. Thus a better method of membrane actuation would be needed. An improved pneumatic method could be achieved by the use of small pistons and linear actuators but ideally a non-pneumatic electrically-driven solution such as electroactive polymers could be employed.

Another limitation exists in the angular resolution of the display. While this may be improved either by using an LCD panel of increased pixel density or a different projection light engine, the optical aberrations from the membrane contribute to the resolution loss. Two methods of improving the optical quality of the deformable membrane beamsplitter could provide promising avenues of research. The first is work on stretchable reflective optical coatings. While much work has been done on many different forms of optical coatings for rigid surfaces, work in elastic optical coatings is lacking. Secondly, the shape of the membrane is not ideal for geometric reflections. It is suspected that by spatially varying the thickness of the membrane, better optical shapes could be achieved for each focal distance.

As mentioned in Chapter 5, a better solution for gaze tracking in NEDs is required. Ideally a gaze tracking method that could provide the needed accuracy, form-factor, long-term usability, and user adaptation can be achieved.



Figure 7.1: A scenario depicting the correct focus support in augmented reality. Both the restaurant and the pamphlet annotations are focused at the appropriate depths. When a viewer focuses on the far buildings (top), the building annotations are clear and sharp while the pamphlet annotations match the pamphlet. When a viewer focuses on the near pamphlet (bottom), the pamphlet annotations are clear and sharp while those indicating the buildings are blurred.

One further concern is that as people age, the effects of presbyopia eliminate the usefulness of a display with large focal range. While I contend that matching the focus of virtual images to the real-world is still useful, presumably the user has some form of correction for the real world, so it would be much better and appeal to a much larger section of the population if dynamic, real-world correction could be achieved in the same device. Initial work has been done by Chakravarthula et al. (2018) showing the potential improvements a real-world correcting display could allow; however that design is limited by a severely diminished FOV. Research into methods for enlarging the FOV for dynamic real-world correcting displays is anticipated.

## **7.2 Conclusion**

From the work presented in this dissertation, I claim that through the use of deformable beam-splitters for see-through near-eye displays, it is possible to create single optical-element, varifocal displays with increased field of view which are better able than previous designs to integrate virtual imagery with the real world in measurable, perceptually advantageous ways. Figure 7.1 depicts a scenario that illustrates correct focus support in AR NEDs.

Despite a few limitations of our system, I believe that providing correct focus cues as well as wide FOV are the most important features of head-mounted displays capable of providing seamless integration of the virtual and the real world. Deformable beamsplitters not only provide the basis for new, improved NED designs, but they can be directly used in perceptual experiments that aim at determining requirements for future systems. As such, this work may facilitate the development of augmented reality technology and contribute to our understanding of how it influences user experience.

This new technology presents a promising direction for future AR displays that require support for focus cues and wide field of view, as well as minimizing compute and power demands, all while maintaining a light form-factor important for prolonged usage.

## REFERENCES

- Akeley, K., Watt, S. J., Girshick, A. R., and Banks, M. S. (2004). A stereo display prototype with multiple focal distances. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 804–813. ACM.
- Akşit, K., Kautz, J., and Luebke, D. (2015). Slim near-eye display using pinhole aperture arrays. *Applied optics*, 54(11):3422–3427.
- Akşit, K., Lopes, W., Kim, J., Shirley, P., and Luebke, D. (2017). Near-Eye Varifocal Augmented Reality Display using See-Through Screens. *ACM Trans. Graph. (SIGGRAPH)*, (6).
- Ando, T. and Shimizu, E. (2001). Head-mounted display using holographic optical element. In *Three-Dimensional Video and Display: Devices and Systems: A Critical Review*, volume 10298, page 1029804. International Society for Optics and Photonics.
- Ando, T., Yamasaki, K., Okamoto, M., Matsumoto, T., and Shimizu, E. (2000). Retinal projection display using holographic optical element. In *Practical Holography XIV and Holographic Materials VI*, volume 3956, pages 211–217. International Society for Optics and Photonics.
- Ando, T., Yamasaki, K., Okamoto, M., and Shimizu, E. (1998). Head-mounted display using a holographic optical element. In *Practical Holography XII*, volume 3293, pages 183–190. International Society for Optics and Photonics.
- Arthur, K. W. (2000). *Effects of Field of View on Performance with Head-Mounted Displays*. PhD Thesis, University of North Carolina at Chapel Hill.
- Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and virtual environments*, 6(4):355–385.
- Baluja, S. and Pomerleau, D. (1994). Non-Intrusive Gaze Tracking Using Artificial Neural Networks. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*, pages 753–760. Morgan-Kaufmann.
- Benko, H., Ofek, E., Zheng, F., and Wilson, A. D. (2015). FoveAR: Combining an Optically See-Through Near-Eye Display with Projector-Based Spatial Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 129–135. ACM.
- Bharadwaj, S. R. and Schor, C. M. (2005). Acceleration characteristics of human ocular accommodation. *Vision Research*, 45(1):17–28.
- Bharadwaj, S. R. and Schor, C. M. (2006). Dynamic control of ocular disaccommodation: First and second-order dynamics. *Vision Research*, 46(6):1019–1037.
- Brewster, S. D. (1849). On new stereoscopes. *Transactions of the Royal Scottish Society of Arts*.

- Brewster, S. D. (1856). *The Stereoscope; Its History, Theory, and Construction, with Its Application to the Fine and Useful Arts and to Education: With Fifty Wood Engravings*. John Murray.
- Brooks, F.P., J. (1999). What's real about virtual reality? *IEEE Computer Graphics and Applications*, 19(6):16–27.
- Burns, P. D. (2000). Slanted-edge MTF for digital camera and scanner analysis. In *Is and Ts Pics Conference*, pages 135–138. SOCIETY FOR IMAGING SCIENCE & TECHNOLOGY.
- Cakmakci, O. and Rolland, J. (2006). Head-worn displays: A review. *Journal of display technology*, 2(3):199–216.
- Campbell, F. W. (1957). The Depth of Field of the Human Eye. *Optica Acta: International Journal of Optics*, 4(4):157–164.
- Campbell, F. W. and Westheimer, G. (1960). Dynamics of accommodation responses of the human eye. *J. Physiol.*, 151(2):285–295.
- Canny, J. (1987). A computational approach to edge detection. In *Readings in Computer Vision*, pages 184–203. Elsevier.
- Chakravarthula, P., Dunn, D., Akşit, K., and Fuchs, H. (2018). FocusAR: Auto-focus Augmented Reality Eyeglasses for both Real World and Virtual Imagery. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2906–2916.
- Cholewiak, S. A., Love, G. S., Srinivasan, P. P., Ng, R., and Banks, M. S. (2017). ChromaBlur: Rendering chromatic eye aberration improves accommodation and realism. *ACM Transactions on Graphics (TOG)*, 36(6):210.
- Ciuffreda, K. J. and Kruger, P. B. (1988). Dynamics of human voluntary accommodation. *American Journal of Optometry and Physiological Optics*, 65(5):365–370.
- Cmglee (2013). English: Density of rod (dotted line) and cone (solid line) photoreceptors along a line passing through the fovea and the blind spot of a human eye vs the angle measured from the fovea, based on 'Foundations of Vision' by Brian A. Wandell.
- Cohen, J. S. (1979). *Three-Dimensional Display of Parallel Sliced Images*. M.S., The University of North Carolina at Chapel Hill, United States – North Carolina.
- Collewyn, H., van der Mark, F., and Jansen, T. C. (1975). Precise recording of human eye movements. *Vision Research*, 15(3):447–IN5.
- Cornsweet, T. N. and Crane, H. D. (1970). Servo-Controlled Infrared Optometer. *JOSA*, 60(4):548–554.
- Cornsweet, T. N. and Crane, H. D. (1973). Accurate two-dimensional eye tracker using first and fourth Purkinje images. *JOSA*, 63(8):921–928.



- Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. In *Electronic Imaging 2004*, pages 36–46. International Society for Optics and Photonics.
- Duane, A. (1922). Studies in monocular and binocular accommodation with their clinical applications. *American Journal of Ophthalmology*, 5(11):865–877.
- Dunn, D. (2019). Required Accuracy of Gaze Tracking for Varifocal Displays. *IEEE VR workshop on Eye Tracking and Vision Augmentation (VisAug)*.
- Dunn, D., Chakravarthula, P., Dong, Q., and Fuchs, H. (2018). Mitigating vergence-accommodation conflict for near-eye displays via deformable beamsplitters. In *Digital Optics for Immersive Displays*, volume 10676, page 106760U. International Society for Optics and Photonics.
- Dunn, D., Tippetts, C., Torell, K., Kellnhofer, P., Akşit, K., Didyk, P., Myszkowski, K., Luebke, D., and Fuchs, H. (2017). Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1322–1331.
- Elliot, J. (1852). On the invention of the stereoscope. *Philosophical Magazine and Journal of Science*, 3(January-June):397.
- Enright, J. T. (1984). Changes in vergence mediated by saccades. *The Journal of Physiology*, 350(1):9–31.
- Erkelens, C. J., Steinman, R. M., and Collewijn, H. (1989). Ocular Vergence Under Natural Conditions. II. Gaze Shifts Between Real Targets Differing in Distance and Direction. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 236(1285):441–465.
- Farrell, J. E., Benson, B., and Haynie, C. R. (1987). Predicting flicker thresholds for video display terminals. In *Proceedings of the SID*, volume 28/4. SID.
- Fattal, D., Peng, Z., Tran, T., Vo, S., Fiorentino, M., Brug, J., and Beausoleil, R. G. (2013). A multi-directional backlight for a wide-angle, glasses-free three-dimensional display. *Nature*, 495(7441):348–351.
- Fincham, E. F. and Walton, J. (1957). The reciprocal actions of accommodation and convergence. *The Journal of Physiology*, 137(3):488–508.
- Findlay, J. M. (1978). Estimates on probability functions: A more virulent PEST. *Perception & Psychophysics*, 23(2):181–185.
- Fry, G. A. (1939). Further Experiments On The Accommodation-Convergence Relationship. *Optometry and Vision Science*, 16(9):325.
- Fuchs, E. (1899). *Text-Book of Ophthalmology*. D. Appleton.
- Fuchs, H., Pizer, S. M., Cohen, J. S., and Brooks, F. P. J. (1980). A three-dimensional display for medical images from slices. *Information processing in medical imaging, Paris, 2-6 July 1979*.

- Fuchs, H., Pizer, S. M., Heinz, E. R., Tsai, L. C., and Bloomberg, S. H. (1982). Adding a True 3-D Display to a Raster Graphics System. *IEEE Computer Graphics and Applications*, 2(7):73–78.
- Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W., and Kasneci, E. (2015). ExCuSe: Robust Pupil Detection in Real-World Scenarios. In Azzopardi, G. and Petkov, N., editors, *Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, pages 39–51. Springer International Publishing.
- Fuhl, W., Santini, T. C., Kübler, T., and Kasneci, E. (2016). ElSe: Ellipse Selection for Robust Pupil Detection in Real-world Environments. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 123–130, New York, NY, USA. ACM.
- Furness, T. A. (1986). The Super Cockpit and its Human Factors Challenges. *Proceedings of the Human Factors Society Annual Meeting*, 30(1):48–52.
- Gao, C., Lin, Y., and Hua, H. (2012). Occlusion capable optical see-through head-mounted display using freeform optics. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 281–282.
- Golay, M. J. E. (1947). Theoretical Consideration in Heat and Infra-Red Detection, with Particular Reference to the Pneumatic Detector. *Review of Scientific Instruments*, 18(5):347–356.
- Groth, H. (2007). Reading Victorian Illusions: Dickens's Haunted Man and Dr. Pepper's "Ghost". *Victorian Studies*, 50(1):43–65.
- Guenter, B., Finch, M., Drucker, S., Tan, D., and Snyder, J. (2012). Foveated 3D graphics. *ACM Transactions on Graphics (TOG)*, 31(6):164.
- Haidinger, W. (1844). Ueber das directe Erkennen des polarisirten Lichts und der Lage der Polarisationssebene. *Annalen der Physik*, 139(9):29–39.
- Hamasaki, T. and Itoh, Y. (2019). Varifocal Occlusion for Optical See-Through Head-Mounted Displays using a Slide Occlusion Mask. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1961–1969.
- Henderson, S. J. and Feiner, S. K. (2011). Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 191–200.
- Heron, G., Charman, W. N., and Schor, C. (2001). Dynamics of the accommodation response to abrupt changes in target vergence as a function of age. *Vision Research*, 41(4):507 – 519.
- Hobgood, W. S. H. (1970). *A Three-Dimensional Computer Graphics Display Using a Varifocal Mirror*. M.S., The University of North Carolina at Chapel Hill, United States – North Carolina.
- Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33–33.

- Hornbeck, L. J. (1997). Digital Light Processing for high-brightness high-resolution applications. In *Projection Displays III*, volume 3013, pages 27–41. International Society for Optics and Photonics.
- Howard, I. P. and Rogers, B. J. (2008). *Seeing in Depth: Volume 1: Basic Mechanics/ Volume 2: Depth Perception 2-Volume Set*. Oxford University Press.
- Hu, X. and Hua, H. (2014). High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Optics express*, 22(11):13896–13903.
- Hua, H. (2017). Enabling focus cues in head-mounted displays. *Proceedings of the IEEE*, 105(5):805–824.
- Hua, H. and Javidi, B. (2014). A 3D integral imaging optical see-through head-mounted display. *Optics express*, 22(11):13484–13491.
- Huang, F.-C., Chen, K., and Wetzstein, G. (2015). The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-eye Light Field Displays with Focus Cues. *ACM Transactions on Graphics (TOG)*, 34(4):60:1–60:12.
- Ilie, A., Low, K.-L., Welch, G., Lastra, A., Fuchs, H., and Cairns, B. (2004). Combining Head-Mounted and Projector-Based Displays for Surgical Training. *Presence*, 13(2):128–145.
- Jang, C., Bang, K., Moon, S., Kim, J., Lee, S., and Lee, B. (2017). Retinal 3D: Augmented Reality Near-eye Display via Pupil-tracked Light Field Projection on Retina. *ACM Transactions on Graphics (TOG)*, 36(6):190:1–190:13.
- Javadi, A.-H., Hakimi, Z., Barati, M., Walsh, V., and Tcheang, L. (2015). SET: A pupil detection method using sinusoidal approximation. *Frontiers in Neuroengineering*, 8.
- Johnson, P. V., Parnell, J. A., Kim, J., Banks, M. S., Love, G. D., et al. (2016a). Assessing Visual Discomfort Using Dynamic Lens and Monovision Displays. In *3D Image Acquisition and Display: Technology, Perception and Applications*, pages TT4A–1. Optical Society of America.
- Johnson, P. V., Parnell, J. A., Kim, J., Saunter, C. D., Love, G. D., and Banks, M. S. (2016b). Dynamic lens and monovision 3D displays to improve viewer comfort. *Opt. Express*, 24(11):11808–11827.
- Kasthurirangan, S., Vilupuru, A. S., and Glasser, A. (2003). Amplitude dependent accommodative dynamics in humans. *Vision Research*, 43(27):2945 – 2956.
- Kaufman, A. E., Bandopadhyay, A., and Shaviv, B. D. (1993). An eye tracking computer user interface. In *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*, pages 120–121.
- Kim, H.-J., Lee, S.-K., Piao, M.-L., Kim, N., and Park, J.-H. (2015). Three-dimensional holographic head mounted display using holographic optical element. In *Consumer Electronics (ICCE), 2015 IEEE International Conference On*, pages 132–133. IEEE.

- Kim, J., Stengel, M., Majercik, A., De Mello, S., Dunn, D., Laine, S., McGuire, M., and Luebke, D. (2019). NVGaze: An Anatomically-Informed Dataset for Low-Latency, Near-Eye Gaze Estimation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '19. ACM.
- King, M. C. and Berry, D. H. (1970). Varifocal Mirror Technique for Video Transmission of Three-Dimensional Images. *Applied Optics*, 9(9):2035–2039.
- Kiyokawa, K. (2007a). An Introduction to Head Mounted Displays for Augmented Reality. *Emerging Technologies of Augmented Reality: Interfaces and Design*, pages 43–63.
- Kiyokawa, K. (2007b). A wide field-of-view head mounted projective display using hyperbolic half-silvered mirrors. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–4. IEEE Computer Society.
- Kiyokawa, K., Billinghamst, M., Campbell, B., and Woods, E. (2003). An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 133. IEEE Computer Society.
- Kollin, J. (1993). A retinal display for virtual-environment applications. *Proceedings of SID International Symposium, Digest Of Technical Papers*, 24:827.
- Konrad, R., Cooper, E. A., and Wetzstein, G. (2016). Novel Optical Configurations for Virtual Reality: Evaluating User Preference and Performance with Focus-tunable and Monovision Near-eye Displays. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'16)*, pages 1211–1220.
- Kotulak, J. C. and Schor, C. M. (1986). The Accommodative Response to Subthreshold Blur and to Perceptual Fading during the Troxler Phenomenon. *Perception*, 15(1):7–15.
- Koulouris, G.-A., Bui, B., Banks, M., and Drettakis, G. (2017). Accommodation and Comfort in Head-Mounted Displays. *ACM Transactions on Graphics*, 36(4):11.
- Kramida, G. (2016). Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1912–1931.
- Kruger, P. B. and Pola, J. (1986). Stimuli for accommodation: Blur, chromatic aberration and size. *Vision Research*, 26(6):957–971.
- Kruger, P. B. and Pola, J. (1987). Dioptric and non-dioptric stimuli for accommodation: Target size alone and with blur and chromatic aberration. *Vision Research*, 27(4):555–567.
- Lambooij, M., Fortuin, M., Heynderickx, I., and IJsselstein, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3):30201–1.
- Landis, C. (1953). *An Annotated Bibliography of the Flicker Fusion Phenomena Covering the Period 1740-1952*. Armed Forces-National Research Council, Vision Committee Secretariat.

- Lanman, D. and Luebke, D. (2013). Near-eye light field displays. *ACM Transactions on Graphics (TOG)*, 32(6):220.
- Lawes, C. P. (2013). Proof Without Words: The Length of a Triangle Median via the Parallelogram Law. *Mathematics Magazine*, 86(2):146–146.
- Lee, S., Hu, X., and Hua, H. (2016). Effects of optical combiner and IPD change for convergence on near-field depth perception in an optical see-through HMD. *IEEE transactions on visualization and computer graphics*, 22(5):1540–1554.
- Li, D., Winfield, D., and Parkhurst, D. J. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, pages 79–79.
- Liang, J., Grimm, B., Goelz, S., and Bille, J. F. (1994). Objective measurement of wave aberrations of the human eye with the use of a Hartmann–Shack wave-front sensor. *JOSA A*, 11(7):1949–1957.
- Lincoln, P., Blate, A., Singh, M., Whitted, T., State, A., Lastra, A., and Fuchs, H. (2016). From Motion to Photons in 80 Microseconds: Towards Minimal Latency for Virtual and Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1367–1376.
- Lincoln, P., Welch, G., Nashel, A., State, A., Ilie, A., and Fuchs, H. (2011). Animatronic shader lamps avatars. *Virtual Reality*, 15(2-3):225–238.
- Lippmann, G. (1908). Epreuves reversibles. photographies integrals. *Comptes-Rendus Academie des Sciences*, 146:446–451.
- Liu, J., Morgens, S.-M., Sumner, R. C., Buschmann, L., Zhang, Y., and Davis, J. (2014). When Does the Hidden Butterfly Not Flicker? In *SIGGRAPH Asia 2014 Technical Briefs*, SA '14, pages 3:1–3:4, New York, NY, USA. ACM.
- Liu, S., Cheng, D., and Hua, H. (2008). An optical see-through head mounted display with addressable focal planes. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium On*, pages 33–42. IEEE.
- Liu, S., Li, Y., Zhou, P., Chen, Q., and Su, Y. (2018). Reverse-mode PSLC multi-plane optical see-through display for AR applications. *Optics Express*, 26(3):3394–3403.
- Llull, P., Bedard, N., Wu, W., Tošić, I., Berkner, K., and Balram, N. (2015). Design and optimization of a near-eye multifocal display system for augmented reality. In *Imaging and Applied Optics 2015 (2015), Paper JTh3A.5*, page JTh3A.5. Optical Society of America.
- Love, G. D., Hoffman, D. M., Hands, P. J., Gao, J., Kirby, A. K., and Banks, M. S. (2009). High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express*, 17(18):15716–15725.

- MacIntyre, B., Bolter, J. D., Moreno, E., and Hannigan, B. (2001). Augmented reality as a new media experience. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 197–206.
- MacKenzie, K. J., Hoffman, D. M., and Watt, S. J. (2010). Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision*, 10(8):22.
- Maimone, A. and Fuchs, H. (2013). Computational augmented reality eyeglasses. In *IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013, Adelaide, Australia, October 1-4, 2013*, pages 29–38. IEEE.
- Maimone, A., Georgiou, A., and Kollin, J. S. (2017). Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics (TOG)*, 36(4):85.
- Maimone, A., Lanman, D., Rathinavel, K., Keller, K., Luebke, D., and Fuchs, H. (2014). Pinlight Displays: Wide Field of View Augmented Reality Eyeglasses Using Defocused Point Light Sources. *ACM Trans. Graph.*, 33(4):89:1–89:11.
- Mansell, J. D., Sinha, S., and Byer, R. L. (2002). Deformable mirror development at Stanford University. In *International Symposium on Optical Science and Technology*, pages 1–12. International Society for Optics and Photonics.
- Martens, T. G. and Ogle, K. N. (1959). Observations on Accommodative Convergence\*: Especially Its Nonlinear Relationships. *American Journal of Ophthalmology*, 47(1, Part 2):455–463.
- Matsuda, N., Fix, A., and Lanman, D. (2017). Focal Surface Displays. *ACM Trans. Graph.*, 36(4):86:1–86:14.
- Maxwell, J., Tong, J., and Schor, C. M. (2010). The first and second order dynamics of accommodative convergence and disparity convergence. *Vision Research*, 50(17):1728–1739.
- McKay, S., Mair, G. M., Mason, S., and Revie, K. (2000). Membrane-mirror-based autostereoscopic display for tele-operation and telepresence applications. In *Electronic Imaging*, pages 198–207. International Society for Optics and Photonics.
- McKay, S., Mason, S., Mair, L. S., Waddell, P., and Fraser, S. M. (1999a). Membrane-mirror-based display for viewing 2D and 3D images. In *Electronic Imaging'99*, pages 144–155. International Society for Optics and Photonics.
- McKay, S., Mason, S., Mair, L. S., Waddell, P., and Fraser, S. M. (1999b). Stereoscopic display using a 1.2-m diameter stretchable membrane mirror. In *Stereoscopic Displays and Virtual Reality Systems VI*, volume 3639, pages 122–132. International Society for Optics and Photonics.
- McQuaide, S. C., Seibel, E. J., Kelly, J. P., Schowengerdt, B. T., and Furness III, T. A. (2003). A retinal scanning display system that produces multiple focal planes with a deformable membrane mirror. *Displays*, 24(2):65–72.

- Mercier, O., Sulai, Y., Mackenzie, K., Zannoli, M., Hillis, J., Nowrouzezahrai, D., and Lanman, D. (2017). Fast Gaze-contingent Optimal Decompositions for Multifocal Displays. *ACM Trans. Graph.*, 36(6):237:1–237:15.
- Mills, P. H., Fuchs, H., and Pizer, S. M. (1984). High-Speed Interaction On A Vibrating-Mirror 3D Display. In *Processing and Display of Three-Dimensional Data II*, volume 0507, pages 93–102. International Society for Optics and Photonics.
- Moon, E., Kim, M., Roh, J., Kim, H., and Hahn, J. (2014). Holographic head-mounted display with RGB light emitting diode light source. *Optics Express*, 22(6):6526–6534.
- Morgan, M. W. (1968). Accommodation and Vergence. *Optometry and Vision Science*, 45(7):417.
- Muirhead, J. C. (1961). Variable Focal Length Mirrors. *Review of Scientific Instruments*, 32(2):210–211.
- Nagahara, H., Yagi, Y., and Yachida, M. (2006). Super wide field of view head mounted display using catadioptrical optics. *Presence*, 15(5):588–598.
- Narain, R., Albert, R. A., Bulbul, A., Ward, G. J., Banks, M. S., and O’Brien, J. F. (2015). Optimal Presentation of Imagery with Focus Cues on Multi-Plane Displays. *ACM Transactions on Graphics*, 34(4):59:1–12.
- Narzt, W., Pomberger, G., Ferscha, A., Kolb, D., Müller, R., Wieghardt, J., Hörtnner, H., and Lindinger, C. (2006). Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187.
- Ogle, K. N. (1932). An Analytical Treatment of the Longitudinal Horopter; Its Measurement and Application to Related Phenomena, Especially to the Relative Size and Shape of the Ocular Images\*. *JOSA*, 22(12):665–728.
- Padmanaban, N., Konrad, R., Stramer, T., Cooper, E. A., and Wetzstein, G. (2017). Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences*, page 201617251.
- Park, S., Spurr, A., and Hilliges, O. (2018a). Deep Pictorial Gaze Estimation. *European Conference on Computer Vision (ECCV)*, 16(1):741–757.
- Park, S., Zhang, X., Bulling, A., and Hilliges, O. (2018b). Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA ’18, pages 21:1–21:10, New York, NY, USA. ACM.
- Percival, A. S. (1910). *The Prescribing of Spectacles*. John Wright.
- Phillips, S., Shirachi, D., and Stark, L. (1972). Analysis of accommodative response times using histogram information. *American Journal of Optometry & Archives of American Academy of Optometry*, 49(5):389–400.

- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., and Williams, S. M. (2001). Types of Eye Movements and Their Functions. *Neuroscience*. 2nd edition.
- Rathinavel, K., Wang, H., Blate, A., and Fuchs, H. (2018). An Extended Depth-of-Field Volumetric Near-Eye Augmented Reality Display. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2857–2866.
- Rawson, E. G. (1968). 3-D Computer-Generated Movies Using a Varifocal Mirror. *Applied Optics*, 7(8):1505–1511.
- Rawson, E. G. (1969). Vibrating Varifocal Mirrors for 3-D Imaging. *IEEE Spectr.*, 6(9):37–43.
- Reulen, J. P. H., Marcus, J. T., Koops, D., de Vries, F. R., Tiesinga, G., Boshuizen, K., and Bos, J. E. (1988). Precise recording of eye movement: The IRIS technique Part 1. *Medical and Biological Engineering and Computing*, 26(1):20–26.
- Roberts, D., Menozzi, A., Cook, J., Sherrill, T., Snarski, S., Russler, P., Clipp, B., Karl, R., Wenger, E., Bennett, M., Mauger, J., Church, W., Towles, H., MacCabe, S., Webb, J., Lupo, J., Frahm, J.-M., Dunn, E., Leslie, C., and Welch, G. (2013). Testing and evaluation of a wearable augmented reality system for natural outdoor environments. In *Head- and Helmet-Mounted Displays XVIII: Design and Applications*, volume 8735, page 87350A. International Society for Optics and Photonics.
- Robinson, D. A. (1963). A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field. *IEEE Transactions on Bio-medical Electronics*, 10(4):137–145.
- Rolland, J. P., Krueger, M. W., and Goon, A. (2000). Multifocal planes head-mounted displays. *Applied Optics*, 39(19):3209–3215.
- Ronchi, L. and Molesini, G. (1975). Depth of Focus in Peripheral Vision. *Ophthalmic Res*, 7(3):152–157.
- Savino, P. J. and Danesh-Meyer, H. V. (2012). *Color Atlas and Synopsis of Clinical Ophthalmology – Wills Eye Institute – Neuro-Ophthalmology*. Lippincott Williams & Wilkins.
- Scholz, J. and Smith, A. N. (2016). Augmented reality: Designing immersive experiences that maximize consumer engagement. *Business Horizons*, 59(2):149–161.
- Schor, C., Wood, I., and Ogawa, J. (1984). Binocular sensory fusion is limited by spatial resolution. *Vision Research*, 24(7):661–665.
- Schor, C. M. (1992). A Dynamic Model of Cross-coupling Between Accommodation and Convergence: Simulations of Step and Frequency Responses. *Optometry and Vision Science*, 69(4):258–269.
- Schor, C. M. and Bharadwaj, S. R. (2005). A pulse-step model of accommodation dynamics in the aging eye. *Vision research*, 45(10):1237–1254.



- Schor, C. M., Lott, L. A., Pope, D., and Graham, A. D. (1999). Saccades reduce latency and increase velocity of ocular accommodation. *Vision Research*, 39(22):3769–3795.
- Schowengerdt, B. T. and Seibel, E. J. (2006). True 3-D scanned voxel displays using single or multiple light sources. *Journal of the Society for Information Display*, 14(2):135–143.
- Schowengerdt, B. T., Seibel, E. J., Kelly, J. P., Silverman, N. L., and Furness, T. A. (2003). Binocular retinal scanning laser display with integrated focus cues for ocular accommodation. In *Stereoscopic Displays and Virtual Reality Systems X*, volume 5006, pages 1–10. International Society for Optics and Photonics.
- Schubert, E. F. (2003). *Light-Emitting Diodes*. Cambridge University Press.
- Seghir, R. and Arscott, S. (2015). Extended PDMS stiffness range for flexible systems. *Sensors and Actuators A: Physical*, 230:33–39.
- Sheard, C. (1934). The Prescription Of Prisms: As Determined by Analyses of Data on Relative Amplitudes of Convergence and Accommodation. *American Journal of Optometry*, 11(10):364–378.
- Shi, L., Huang, F.-C., Lopes, W., Matusik, W., and Luebke, D. (2017). Near-eye Light Field Holographic Rendering with Spherical Waves for Wide Field of View Interactive 3D Computer Graphics.
- Shibata, T., Kim, J., Hoffman, D. M., and Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of vision*, 11(8):11–11.
- Sisodia, A., Riser, A., and Rogers, J. R. (2005). Design of an advanced helmet mounted display (AHMD). In *Defense and Security*, pages 304–315. International Society for Optics and Photonics.
- Spencer, G. and Murty, M. (1962). General ray-tracing procedure. *JOSA*, 52(6):672–676.
- staff, B. (2014). Medical gallery of Blausen Medical 2014. *WikiJournal of Medicine*, 1(2).
- Sutherland, I. E. (1965). The ultimate display. *Multimedia: From Wagner to virtual reality*, pages 506–508.
- Sutherland, I. E. (1968). A Head-mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, Part I), pages 757–764, New York, NY, USA. ACM.
- Sweeney, L. E., Seidel, D., Day, M., and Gray, L. S. (2014). Quantifying interactions between accommodation and vergence in a binocularly normal population. *Vision Research*, 105:121–129.
- Taylor, M. M. and Creelman, C. D. (1967). PEST: Efficient Estimates on Probability Functions. *The Journal of the Acoustical Society of America*, 41(4A):782–787.

- Temple, S. E., McGregor, J. E., Miles, C., Graham, L., Miller, J., Buck, J., Scott-Samuel, N. E., and Roberts, N. W. (2015). Perceiving polarization with the naked eye: Characterization of human polarization sensitivity. *Proceedings of the Royal Society B: Biological Sciences*, 282(1811):20150338.
- Tew, A. I. (1997). Simulation results for an innovative point-of-regard sensor using neural networks. *Neural Computing & Applications*, 5(4):230–237.
- Traub, A. C. (1967). Stereoscopic Display Using Rapid Varifocal Mirror Oscillations. *Applied Optics*, 6(6):1085–1087.
- Treloar, L. R. G. (1944). The Elasticity of a Network of Long-Chain Molecules. II. *Rubber Chemistry and Technology*, 17(2):296–302.
- Tsurutani, K., Naruse, K., Oshima, K., Uehara, S., Sato, Y., Inoguchi, K., Otsuka, K., Wakemoto, H., Kurashige, M., Sato, O., et al. (2017). Optical Attachment to Measure Both Eye-Box/FOV Characteristics for AR/VR Eyewear Displays. In *SID Symposium Digest of Technical Papers*, volume 48, pages 954–957. Wiley Online Library.
- Vera, L., Gimeno, J., Coma, I., and Fernández, M. (2011). Augmented Mirror: Interactive Augmented Reality System Based on Kinect. In Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., and Winckler, M., editors, *Human-Computer Interaction – INTERACT 2011*, Lecture Notes in Computer Science, pages 483–486. Springer Berlin Heidelberg.
- von Waldkirch, M. (2004). *Retinal Projection Displays for Accommodation-Insensitive Viewing*. Doctoral Thesis, ETH Zurich.
- von Waldkirch, M., Lukowicz, P., and Tröster, G. (2003a). Defocusing simulations on a retinal scanning display for quasi accommodation-free viewing. *Optics express*, 11(24):3220–3233.
- von Waldkirch, M., Lukowicz, P., and Tröster, G. (2003b). LCD-based coherent wearable projection display for quasi-accommodation-free imaging. *Optics communications*, 217(1-6):133–140.
- von Waldkirch, M., Lukowicz, P., and Tröster, G. (2005). Oscillating fluid lens in coherent retinal projection displays for extending depth of focus. *Optics communications*, 253(4-6):407–418.
- Wall, F. T. (1942). Statistical Thermodynamics of Rubber. II. *The Journal of Chemical Physics*, 10(7):485–488.
- Wang, B. and Ciuffreda, K. J. (2004). Depth-of-focus of the human eye in the near retinal periphery. *Vision Research*, 44(11):1115 – 1125.
- Wang, B. and Ciuffreda, K. J. (2006). Depth-of-Focus of the Human Eye: Theory and Clinical Implications. *Survey of Ophthalmology*, 51(1):75 – 85.
- Wang, X., Love, P. E. D., Kim, M. J., and Wang, W. (2014). Mutual awareness in collaborative design: An Augmented Reality integrated telepresence system. *Computers in Industry*, 65(2):314–324.

- Watson, A. B. (2013). High Frame Rates and Human Vision: A View through the Window of Visibility. *SMPTE Motion Imaging Journal*, 122(2):18–32.
- Werber, A. and Zappe, H. (2008). Tunable pneumatic microoptics. *Journal of Microelectromechanical Systems*, 17(5):1218–1227.
- Westheimer, G. (1966). The maxwellian view. *Vision research*, 6(11-12):669–682.
- Weymouth, F. W. (1958). Visual sensory units and the minimal angle of resolution. *American journal of ophthalmology*, 46(1):102–113.
- Wheatstone, C. (1838). XVIII. Contributions to the physiology of vision. —Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128:371–394.
- Yamazaki, S., Mochimaru, M., and Kanade, T. (2011). Simultaneous Self-Calibration of a Projector and a Camera Using Structured Light. In *Proc. Projector Camera Systems*, pages 67–74.
- Yuuki, A., Itoga, K., and Satake, T. (2012). A new Maxwellian view display for trouble-free accommodation. *Journal of the Society for Information Display*, 20(10):581–588.
- Zannoli, M., Love, G. D., Narain, R., and Banks, M. S. (2016). Blur and the perception of depth at occlusions. *Journal of Vision*, 16(6):17–17.
- Zone, R. (2007). *Stereoscopic Cinema and the Origins of 3-D Film, 1838-1952*. University Press of Kentucky.