Bangor University

**DOCTOR OF PHILOSOPHY**

# The importance of correct focus cues in 3d stereoscopic imagery.

Anantha Krishnan, Anantha

*Award date:*
2023

*Awarding institution:*
Bangor University

Link to publication

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

**Take down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 30. Apr. 2024

# The importance of correct focus cues in 3d stereoscopic imagery.

Anantha Krishnan, MSc

Submitted for the degree of Doctor of Philosophy

2022

School of Human and Behavioural Sciences
Bangor University

‘I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

I confirm that I am submitting this work with the agreement of my Supervisor(s).’

---

‘Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw’r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o’r blaen ar gyfer unrhyw radd, ac nid yw’n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

Rwy'n cadarnhau fy mod yn cyflwyno'r gwaith hwn gyda chytundeb fy Ngoruchwyliwr (Goruchwylwyr)

Acknowledgement

Firstly, I would like to express my sincere gratitude to my advisor Dr Simon Watt for his continuous support not only with research but life outside research. I sincerely appreciate him for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would also like to thank my second-supervisor Dr Rafał Mantiuk, for assisting me in my research. I learnt organisation, communication, and group management from him. In short how to efficiently lead a research group.

Besides my advisors, I would like to thank my thesis committee chair: Dr Kami Koldewyn for her insightful comments and encouragement through tough COVID-19 times, but also for the hard question which helped me to widen my research from various perspectives.

I would also like to thank my fellow lab colleagues, Michela Paroli, Pierre-Arthur Suray, Molly Hewitt, for helping me out wherever I needed a third hand.

Since my PhD was part of a European Horizon 2020 project (ITN RealVision), I had the opportunity to work and spend leisurely time with a few wonderful people: Waqas Ellahi, Ali Ak, Milan Stepanov, Abhishek Goswami, Fangcheng Zhong, Alie Hexley, Krzysztof Wolski, Muhammad Shahzeb Khan Gul, Jingyu Liu, Muhammad Umair Mukati, Sarvesh R. Agrawal, Akshay Jindal, and Randy Frans Fela.

Last but not the least, I would like to thank my family: my parents and to my brother for supporting me spiritually throughout writing this thesis and my life in general.

# Table of content

**Chapter 3**



**Chapter 4**



**Chapter 5**

## Table of Figures

**Chapter 1**



**Chapter 2**

**Chapter 3**



**Chapter 4**

**Chapter 5**

## Abstract

Stereoscopic 3-D display technologies aim to provide a compelling, realistic sense of 3-D depth. Theoretically, this could be achieved by having a system that can project the same light rays coming from the real world via a display. Recent advances in rendering, display, and image acquisition have made it possible to present almost all depth cues accurately. Focus cues, including the eye's focusing response to image blur (accommodation) and the pattern of retinal blur arising from objects at varying distances (retinal blur gradient), remain challenging to replicate accurately, however. In conventional stereoscopic displays, focus cues are incorrect, as the depth is specified by the screen surface and not by the position of the content in the scene. This issue causes two problems: firstly, the unnatural pattern of oculomotor responses elicited when viewing the 3-D scene (inaccurate accommodation signals), and the appearance (inaccurate retinal blur gradient) may signal the unnaturalness of depth in the scene (Hibbard et al., 2017). Secondly, the inconsistency between depth specified by vergence from disparity and focus cues results in a conflict commonly known as vergence-accommodation conflict. The visual system's effort to resolve this conflict causes visual discomfort and reduced stereoacuity (Watt et al., 2005 a; Hoffman et al., 2008; Shibata et al., 2011). There is an increasing demand for stereoscopic 3-D displays that can elicit realistic depth and have minimal to no user issues. Several technologies have tried to address the issue with incorrect focus cues, but all are computationally intensive, require specialist hardware, and compromise image quality. Hence, the value of presenting focus cues accurately needs to be fully understood. In Chapter 2, we measured the vergence distances (zone of good stereo depth perception) around the screen, where stereoscopic depth perception remains effective despite incorrect focus cues. To understand whether all users would benefit similarly from correct focus cues, we specifically examined whether age and age-related changes in the ability to accommodate predicted individuals' tolerance to vergence -accommodation conflict. We used a fixed-viewpoint volumetric display (specifically Multiple-focal plane display, Mackenzie et al., 2010) to present stimulus with varying vergence-accommodation conflict to see at what conflict value individuals' stereoacuity drop to a criterion level. We did not

find any predictive effect of age or people’s ability to accommodate on their tolerance to vergence-accommodation-induced degradation in depth perception. However, we found other visual factors, such as phroia, to predict the symmetry of the centre of the zone (similar to Shibata et al., 2011). And guidelines for stereo 3-D content optimisation. In Chapter 3, we explored whether depth appears more realistic when focus cues are correct. We used a fixed-viewpoint volumetric display (specifically Multiple-focal plane display, Mackenzie et al., 2010) to present stimulus with and without natural variations in focus cues. Participants made a two-interval, forced-choice response indicating which interval contained: Experiment:1, the largest depth separation or Experiment:2 which interval contained the most tangible, solid and real depth separation. The results provide strong evidence that focus cues are important for presenting scenes with realistic depth, and, that individuals can separate the quantitative and qualitative aspects of stereoscopic depth. In Chapter 4, we evaluated whether the effect of focus cues on realism is detectable in high-fidelity images and if depth-of-field (DoF) rendering can be used to substitute any effects of focus cues (using a similar method in Chapter 3). The results show that focus cues do increase depth realism in a 3-D scene for high-fidelity images, and the effects of focus cues cannot be substituted with DoF rendering techniques. Our study has shown that we can optimise certain aspects of 3-D content to suit an individual's tolerance to vergence-accommodation conflict by utilising visual factors, such as their phoria. However, the technical constrains are difficult to achieve if we are to get focus cues correct. So, depending on the application's requirements, a decision must be made between stimuli with tolerable conflict and realistic depth.

# Chapter 1

# 1 General introduction

## 1. Introduction

Stereoscopic 3-D imagery is used to provide a percept of three-dimensional (3-D) depth structure, which is otherwise absent in a painting, rendered computer image or photograph. This is achieved by presenting different images to the two eyes, using one of a number of technologies, to simulate the natural binocular disparities—differences between the positions of corresponding points in the two eye's images—that occur when viewing a real scene with two eyes separated laterally in the head (Banks et al., 2012). The purpose of much stereoscopic 3-D imagery is not only to create a strong sense of 3-D structure but also to produce a sense of realness or naturalness to the sense of depth, that we associate with viewing a real-world scene (referred to as depth realism, Hibbard et al., 2017).

## 1.1 Applications of stereoscopic 3-D imagery

A comprehensive review of the various applications of stereoscopic 3-D imagery is beyond the scope of this thesis. A brief summary is given below.

Stereoscopic 3-D imagery has been known since the nineteenth century. Sir Charles Wheatstone demonstrated a stereoscope, where the image of the same scene is presented from two vantage points to each eye separately, using mirrors, and individuals fuse the images, and are able to see stereo depth in the scene (Wheatstone, 1838). Since then, this technique has only been adopted into niche areas, due to its limited application potential. In the last decade the popularity of stereoscopic 3-D displays has increased exponentially, as the display and image projecting technologies (especially affordable VR headsets), and software and content, have evolved and matured, and new fields have emerged which would benefit from 3-D technology.

The relatively economical availability of miniaturized circuit chips and hardware with high computing power has resulted in lifting the long-standing barriers associated with stereo 3-D in medical applications. Today stereoscopic 3-D imagery also has direct applications in the field of education, 3-D modelling and designing, and computer vision. Interactive stereo 3-D displays are used in the development of tools for neurosurgical education (Henn et al., 2002). Due to high cost associated with laboratory management and administration, cadavers are getting replaced with clinically representative 3-D rendered stereoscopic models, which can be accessed using modern head-mounted displays (HMD's) (Simpson, 2014; Brewer-Deluce et al., 2021). Many institutions that specialize in cardiology care services are taking advantage of stereoscopic 3-D for demonstration, pre-procedural, procedural, and post-procedural planning and visualization (Silva et al., 2018). Stereoscopic head mounted displays (HMD's) have shown to help with neurorehabilitation, in patients with stroke (Laver et al., 2015; Peroz-Marcos et al., 2017), allowing practitioners to prescribe higher dosage (increased therapy time) of functional therapy sessions compared to conventional methods (Laver et al., 2015).

Historically scientific visualization and demonstration of complex processes and structures was been done on 2-D media, including TV, books, digital and OHP light projectors. But they are limited by the number of axes that can be used to present a 3-D object and require complicated diagrams from multiple viewpoints to explain the system. However, stereoscopic 3-D displays enable us to present 3-D structures in a more intuitive manner that resembles the real counterpart more faithfully in its 3-D structure and geometry (Ravanagh et al., 2017). Computer vision is the latest arena where stereo 3-D technologies are readily used. In computer vision, the machine is trained to interpret and understand a real-world scene and its entities using clever algorithms (Ballar & Brown, 2022). Stereoscopic 3-D based image acquisition techniques have been shown to help in the improvement of computer vision application concerned with 3-D reconstruction, object detection, segmentation, etc. by providing information of scene depth, colour variation, contrast, and other complex details of the real-world scene (Wu et al., 2017). Stereo 3-D is now also common in entertainment applications, including 3-D cinema and gaming.

The current momentum of research in the field of stereoscopic 3-D has been in the development of displays & rendering techniques that can produce images or scenes that provide a realistic sense of 3-D (Banks et al., 2016). The basic idea is to have a system that can in theory project the same light rays coming from the real world via a display.

## 1.2 Depth Perception in humans

Humans see the world using information gathered from light rays that enter the eyes. The properties of the light rays falling on the retina are influenced by the luminance, distance and other properties of the objects in the scene. Perception of depth in a scene is influenced by the properties of the retinal image, and associated extra retinal signals, collectively known depth cues (Goldstein, 1989; Palmer, 1999; Banks et al., 2016). According to Banks et al., (2016), these depth cues can be classified into three categories based on how they are derived:  perspective projection-based depth cues, light transport and reflection-based depth cues, and triangulation-based depth cues (Banks et al., 2016).

*a. Perspective projection-based depth cues*

These depth cues are governed by the geometry of the projection of the world on the retina, and how the visual system interprets it (Gombrich, 1969; Palmer, 1999; Rogers & Howard, 2012). This includes the following cues to depth.

*Linear perspective*: Parallel lines that recede into distance appear to converge and are readily used by artists to simulate a sense of depth in a scenery (Pizlo & Scheessele, 1998; Palmer, 1999; Rogers & Howard, 2012) (Figure 1.1).

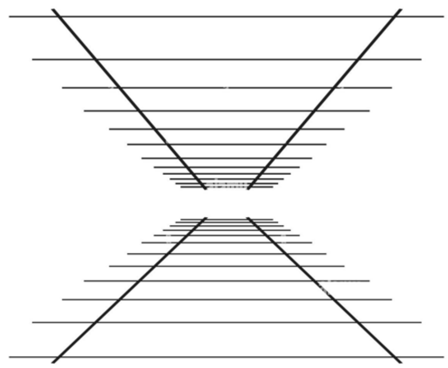


**Figure 1.1:** Shows an example of linear perspective. The two parallel lines appear to converge towards the horizon, this gives a cue to increasing distance towards the horizon.

*Relative size*: For two objects that are the same size, an object that is nearer to the eyes occupy a larger area on the retina (Palmer, 1999; Rogers & Howard, 2012).

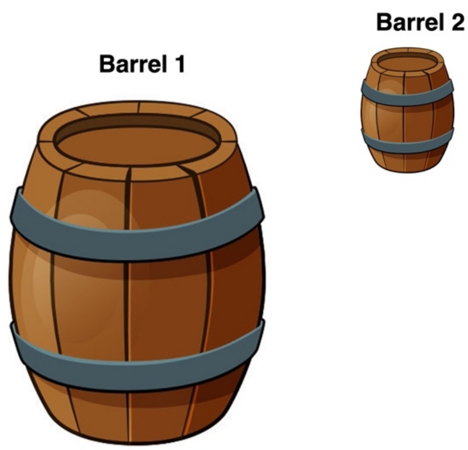



**Figure 1.2:** This figure shows an example of relative size as depth cue. Barrel 1 is bigger than barrel 2, and this gives a sense that barrel 1 is closer to the viewer than barrel 2.

*Texture gradient*: Most surfaces found in real world such as an open field, or a field of rice/ wheat, roads., etc have texture. As distance increases the texture elements are smaller in the retinal image. This feature is often replicated to rendered depth in a 3-D scene (Figure 1.3) (Gombrich, 1969; Palmer, 1999; Rogers & Howard, 2012).

**Figure 1.3:** And example of perspective depth cue from texture. The texture on this stone pavement becomes more dense with distance.

*b. Light transport and reflection-based depth cues*

In the real-world light rays may undergo reflection, refraction, diffusion, scattering (splitting white light into its constituent wavelengths), and specular reflection on lustrous surfaces, before entering the eye. These interactions of light rays in the scene produces various depth cues discussed below.

*Shading*: Entities in the scene that are nearer to the source of light are more illuminated than those farther away, and the brain can infer depth from the relationship between the light source and the objects by shadows cast by them in the scene. Shading also conveys information about the shape of objects. Different regions on a curved surface, for example, reflect different amounts of light to the eye due to their different surface orientation with respect to the eye, creating gradual variations in shading that depend on the 3-D surface structure.

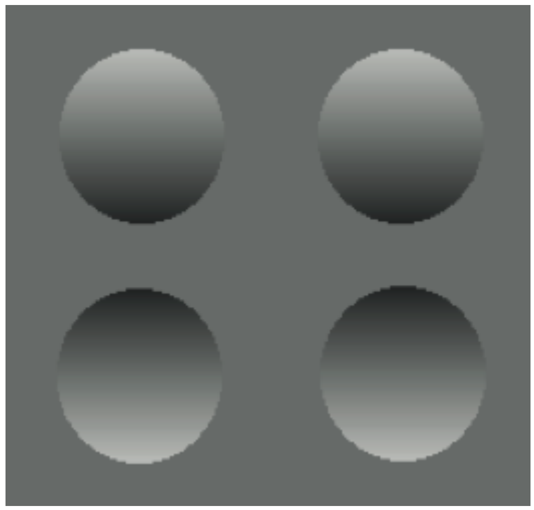


**Figure 1.4:** Shading as depth cue. The direction of the shadow in the circles gives an illusion of Top: convex (bump) and Bottom: concave (dip).

*Occlusions:* Objects in-front-of other objects obstructs the view of the object behind it, giving a cue of the object occluding the other being closer.

**Figure 1.5:** Occlusion as a depth cue: The shapes occluding the other seems to be nearer to the eye than the shapes being occluded, this natural phenomenon is often used to simulate depth in a 3-D scene.

*Atmospheric attenuation*: The fine details and the contrast of objects farther away are not visible and tend to be hazy. This is due to scattering of light by dust and moisture in the air, and causes high frequency details such as the sharp boundaries of objects are not resolvable at large distances (Figure 1.6; O'Shea et al., 1994; Palmer, 1999). Shorter wavelengths of light tend to scatter more, so objects at larger distances have a more bluish tint, and so this serves as a cue to distance (Figure 1.6) (Gao et al., 2012).

**Figure 1.6:** Atmospheric attenuation as a depth cue: As the distance in the scene increases the mountains and the trees loose it's sharpness, and the scene appears to be enveloped in a blueish tint.

*c. <u>Triangulation based cues</u>*

The brain is able to derive depth information when dealt with two slightly different views of the same scene either by virtue of two different eye's images (Figure 1.7.A) (binocular disparity), light rays passing through different vantage points of the pupil (Figure 1.7.B) (retinal blur) or by a temporal variation in view caused by motion (motion parallax) (Palmer, 1999; Banks et al., 2016).

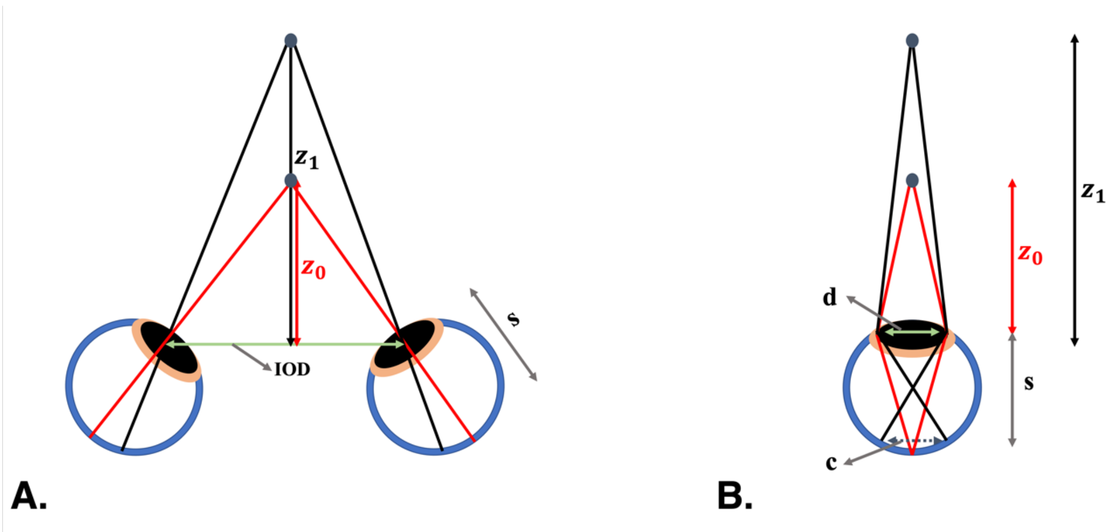



**Figure 1.7**: Geometry of Disparity and Blur: A. In this depiction, the left and right eyes are separated by a distance known as the interocular distance (IOD) and are both focused on a common point at distance $z_0$. An object positioned at a different distance, $z_1$, projects its image onto a parafoveal area of the retinas. Disparity, in degrees or arcseconds, is computed as the angular difference between the projections of $z_1$ onto the two retinas, as shown in (A)**.** B. When an eye is fixated at a distance $z_0$ objects located at other distances will appear blurred on the retina. The object at distance $z_1$ experiences blur, creating a circular region with a diameter represented as "c". The boundaries of this blur circle can be geometrically likened to the projections of $z_1$ on the two retinas as in (A).

### 1.2.2 Retinal and associated extra-retinal cues

All of the above cues are retinal-image-based cues, in that they provide information via the retinal image. There are also so-called extra-retinal cues, which provide depth information via non-retinal mechanisms. For instance, binocular disparity is accompanied by vergence—equal-and-opposite (disconjugate) eye movements—made to align the foveas of both the eyes with the point of interest in the scene (Palmer, 1999). Vergence and binocular disparity forms the basis for stereoscopic 3-D perception in humans, with vergence providing an estimate of the distance to fixation, which is then used to interpret disparities (Howard and Rogers, 2002). Together they are referred to as stereo cues (Banks et al., 2016). The extraretinal cues associated with motion parallax include kinaesthesia (the ability of human body to be aware of the location of parts of the body in space by virtue of proprioceptive signals from the muscle and joints) and vestibular signals (vestibular receptors in the inner-ear sense linear and rotational motion, and head tilts enabling the human body to maintain balance while in rest and in motion). Together they are called parallax cues (Howard and Rogers, 2002). The extraretinal signal associated with retinal blur is accommodation—the focusing ability of human eyes by changing the shape of the crystalline lens via the ciliary muscles. It is unclear whether there is an afferent signal from the ciliary muscles to signal accommodative state. However, it appears the brain can monitor the outgoing signal sent to the muscle (efference copy; Howard and Rogers, 2002). Together, accommodation and retinal blur are known as focus cues. It can be argued that a fundamental principle of creating a stereoscopic 3-D display is to try maintain the normal, correct relationship between retinal and extraretinal cues (Banks et al., 2016).

## 1.3 Getting stereo and parallax cues correct

Parallax and stereo cues are dependent on the point from which the scene is viewed, and are thus based on the same geometry (Koenderink, 1986; Rogers and Graham, 1982). Extensive development of methods to separate the two eye's images, high display refresh rates, and accurate near-real-time gaze- and head-tracking has over the years enabled increasingly accurate presentation of stereo and parallax cues (Pastoor & Wopking, 1997; Suyama et al., 2000; Lin & Hua, 2009; Love et al., 2009).

Stereo cues (binocular disparity and vergence) can be presented relatively accurately with current technologies. Displays that deliver stereoscopic images use various methods to present a pair of same images from two slightly different vantage points/ perspective using two separate displays, or a single display and either passive filters (e.g. red-blue anaglyph, Pstoor & Wopking, 1997), polarizing filters (Sexton & Surman, 1999), or active shutter glasses (e.g. liquid crystal shutter glasses). The two eye's images are fused to give a sensation of stereoscopic depth to the users (Gernsheim & Gernsheim,1969).

The image separation is achieved by either an overlapping optical path (both images are presented on a single screen and each image is encoded with separate wavelength or polarization) or non-overlapping optical path (images for both the eyes are presented on separate screens or on two different positions of a screen) solutions (Akeley et al., 2004; Love et al., 2009; MacKenzie et al., 2010). Anaglyph and polarization-based stereoscopy are examples of overlapping optical path displays. In a polarization-based stereoscope, the images for left and right eye are produced using different orientations of polarization on a single screen and users wear goggles with polarized filters to separate out desired images to the intended eye (Borel & Doyen, 2013). In an anaglyph bases display, instead of differential polarization, image separation is achieved by two separate colours (wavelengths) (Jorke et al., 2009; Woods & Harris, 2010; Simon & Jorke, 2011; Borel & Doyen, 2013). Stereoscopes with overlapping optical paths require only a single display, and (usually) no optical components. However, they can suffer from 'crosstalk' (where one eye can partially see the other eye's image), and distorted colour appearance, etc, (Formankiewicz & Mollon, 2009; Jack, 2012).

Non-overlapping optical path stereoscopic displays are arguably the most preferred method for stereoscopic presentation in research settings as they guarantee no crosstalk's between the two images. Head mounted displays (HMD's) and mirror stereoscopes are examples for such displays (Wheatstone, 1838; Howard and Rogers, 2002).

Parallax cues can faithfully be produced in virtual reality devices by tracking the head position of the user and updating the graphics accordingly as the head position and orientation changes (Thatte et al., 2016). Recent developments in position and orientation trackers (gyroscopes, motion sensors, etc.), fast image-based rendering pipelines and 3-D data representation methods (e.g. depth augmented stereo panoramas [DASP]) have enabled presentation of fairly accurate parallax cues in near real-time (Thatte et al., 2016; Banks et al., 2016).

## 1.4 Sensory integration aspect of depth perception

Our visual system does not rely on only one depth cue, but it derives depth information from a variety of depth cues simultaneously (Buckley & Frisby, 1993; Backus & Banks, 1999; Jacobs, 1999; Knill & Saunders, 2003; Hillis et al., 2004). All the depth cues are combined optimally, i.e., the resulting depth information is highly precise. It has been shown that the optimal way to combine multiple depth cues is as a weighted sum of the estimates from each cue, assuming that the cues' measurement noises are Gaussian distributed and independent of each other, and all depths have the same probability (Cochran, 1937; Ghahramani et al., 1997, Banks et al., 2016):

$$\hat{D} = \sum_i w_i D_i \qquad (1)$$

where the weights are determined by how reliable each cue is

$$w_i = \sigma_i^{-2} \mid \sum_j \sigma_j^{-2} \qquad (2)$$

where $\hat{D}$ is relative depth in the scene specified by cue $i$ , and $\sigma_i$ is standard deviation of the estimate from a given cue. In equation 2 we can see that weights $w_i$ of the cues are proportional to the normalized inverse variances (sigma squared); hence, more weight is allocated to less variable/more reliable cues (Cochran, 1937; Ghahramani et al., 1997; Wolpert & Jordan, 1997; Jacobs, 1999; Backus & Banks, 1999; Ernst & Banks, 2002).

The variance is lower for combined depth estimates than for any lone depth cue. The brain can thus estimate depth with greater precision when information from several depth cues is combined, rather than relying on a single depth cue. It can also do so flexibly, across changing cue reliabilities (which change with distance, surface orientation etc.; Knill and Saunders, 2003; Hillis et al., 2004). There is evidence supporting that when combining sensory signals, the reliability of the cue is considered (Backus & Banks, 1999; Buckley & frisby, 1993; Jacobs, 1999; Körding & Wolpert, 2004; van Beers et al., 1998). Several studies have evaluated the quantitative predictions of this integration model by measuring reliabilities of each cue independently and then using these to predict performance in multiple cue condition, and found near-optimal integration of cues to 3-D scene structure (Ernst & Banks, 2002; Gepstein & Banks, 2003; Hillis et al., 2004; Knill & Saunders, 2003; Landy & Kojima, 2001).

When viewing real world scenes, all the cues specify the same depth, and so changes in cue weights do not change the bias (accuracy) in perception of 3-D structure (this is however only true if all the depth estimates are assumed to be unbiased, which is not the case in few natural scenarios). But, when viewing conventional 3-D displays the scene is presented at a single focal distance. This means that while most cues can be presented correctly (albeit with head tracking in the case of motion parallax or a stereoscopic presentation system for disparity), focus cues are incorrect because they specify the display surface, not the depth structure of the depicted scene (see below). If focus cues are weighted significantly, they may therefore lead to biases in depth perception in stereo 3-D. Focus cues present a particular challenge for stereo 3-D displays because they are difficult to present correctly.

## 1.5 Focus cues (Accommodation and retinal blur)

### 1.5.1 Anatomy of focus cues

Human eyes, much like a camera, have an optical component (a convex lens formed by the tear film, cornea and crystalline lens), a photo sensitive element (the retina) and an entrance aperture through which light rays pass (the pupil) (Remington, 2012). Humans perceive images when light reflected from the surroundings falls on the photosensitive receptors (rods and cones) that comprise the retina. The retina is anatomically divided into fovea and parafoveal regions based on the concentration of the photoreceptors; the fovea has the highest effective spatial resolution due in part to the concentration of photoreceptors (mainly cones), but also to due to photoreceptor outputs being spatially pooled in the retinal periphery (Remington, 2012). When we fixate at a point in the world our eye changes its focal power (accommodates) to achieve a sharp foveal retinal image. This is achieved by changing the shape of the crystalline lens, by actuating the ciliary muscles to relax and contract the zonular fibres that support the lens structure. A consequence of this optical arrangement is that, again much like a camera, points in the scene that are nearer or farther than the distance the eye is accommodated to are subject to defocus blur. The overall pattern of blur caused by differential defocus error is referred to as the retinal blur gradient. Greater defocus, caused by image points being at a greater distance (nearer or farther) than the current accommodative distance, results in greater blur. Note however that the amount of blur for a given defocus, is not fixed but depends on the size of the pupil (in the same way that depth of field in photography depends on the camera aperture).

This means that while retinal blur gradient conveys information about relative distance absolute distance may not be obtainable (see below). In principle, knowledge of the current accommodative state—the distance the eye is focused at—can provide an absolute estimate of distance. There is not thought to be an afferent signal from the ciliary muscles but the efferent signal could be a depth cue, given that the magnitude of the accommodation response depends directly on the distance to the fixated object.

### 1.5.2 Contribution of focus cues to visual depth perception

When viewing a real scene, retinal blur varies consistently with variation in the depth in the scene. The objects to which the eye is focused have the sharpest retinal image and nearer and farther objects are blurred. As described above, retinal blur is to a first approximation an ordinal depth cue, and so has traditionally been considered a weak source of depth information (Mather et al., 2006). Furthermore, again to a first approximation, defocus blur is unsigned (i.e., the same magnitude of blur is present in the scene for points nearer or farther than the point of focus (Mather, 2006; Mather and Smith, 2000). There is evidence, however, that the sign of defocus blur can be recovered from a combination of chromatic aberration and higher-order aberrations (Fincham, 1951; Kruger et al., 1993; Aggarwala et al., 1995; Lee et al., 1999; Fernandez & Artal, 2005; Chen et al., 2006; Chin et al., 2009a; Cholewiak et al., 2017), as well as contrast changes caused by continuous microfluctuations in accommodation (Campbell et al., 1959; Charman & Tucker, 1978; Kotulak & Schor, 1986; MacKenzie et al., 2010). Consistent with this, Nguyen, et al. (2005) found that observers could perceptually discriminate the sign of a step in depth defined only by retinal blur (they used real objects with an edge carefully constructed so there were no size cues), though performance worsened considerably under monochromatic illumination, which eliminates chromatic aberrations. More recently, Cholewiak et al. (2017) developed a computer graphics technique to render blur with near-correct chromatic aberration. Critically, the rendering takes into account the eye’s optics, creating an image on the screen that, when passed through the eye’s optics, creates chromatic aberration in the retinal image (via convolution of the image with a model of the eye’s optics). Cholewiak et al. (2017) found this stimulus caused an accommodation response in the appropriate direction (it also slightly increased judgements of realism of the depicted scene).

Other studies point to a role of blur in the magnitude of perceived depth. Watt et al. (2005) measured the direct contribution of focus cues to perceived slant by independently varying the physical slant of a display surface and the slant of a simulated surface specified by binocular disparity alone, or texture (perspective) alone (monocular viewing). In the binocular condition, the slant of the screen did not affect estimates of slant, but in the monocular condition the slant estimates were affected by the screen surface slant, indicating that focus cues contributed to perceived slant when other cues are less reliable (as predicted by cue integration; see earlier). More recently, Held et al. (2012) showed that depth information from blur is more reliable than depth from disparity for image points that lie nearer or farther than fixation—i.e. away from the horopter—where disparity sensitivity becomes quite poor. Moreover, because blur relies on fewer assumptions about the world than cues such a perspective, they suggested it may be more reliable than those cues, and so may play a more significant role in depth perception than has previously been realised, particularly in complex scenes with multiple objects, as opposed to single, isolated objects typical of vision science experiments (Held et al., 2012).

Although accommodation can also in principle provide direct information about distance to points in the scene (see earlier), on its own it has been shown to provide only a noisy distance cue (Fisher & Ciuffreda, 1988; Künnapas, 1968; Mon-Williams & Tresilian, 2000). Accommodation can also affect the depth in a scene indirectly through the process of disparity scaling, however. The binocular disparity ($\delta$) created by two points in space is related to the viewing distance as per the following equation:

$$\boldsymbol{\delta \approx \frac{I\Delta D}{D^2}} \qquad (3)$$

where $D$ is the viewing distance, $\Delta D$ is the distance between two points in depth, and $I$ is the inter-ocular distance (Howard & Rogers, 2002). The non-linear dependence of disparity on viewing distance means that to determine the depth in the world between two points, the disparities must be 'scaled' by an estimate of the viewing distance. Both vergence, and the horizontal gradient of vertical disparity contribute to this estimate (Rogers & Bradshaw, 1993,1995). However, it has also been shown that the stimulus to accommodation affects disparity scaling (Watt et al., 2005; Hoffman et al., 2008; Vienne et al., 2018), suggesting that accommodation-specified distance plays a role in perception of depth.

Another way in which focus cues have been shown to affect perception of 3-D structure is by effects of manipulating the 'global' blur gradient of an image, which can dramatically alter the overall sense of scale of the depicted scene. This is often known as tilt-shift miniaturisation, where a full-scale scene is made to look miniature by exaggerating the blur gradient in the image (either in software, or by using so-called tilt-shift cameras, that alter the normal relationship between the lens and the film plane (Held et al., 2010; Laforet, 2007; Vishwanath & Blaser, 2010). This effect can also be reversed (by reducing the blur gradient) to make small scenes look large, and is used in filmography, where scale models can be made to appear full size (Fielding, 1985). The tilt-shift effect demonstrates that under some circumstances blur can have very large effects on perception of 3-D structure.

Finally, there is evidence that blur can affect influence only the magnitude of perceived depth, but also the quality of depth percepts. Viewing real 3d scenes results in a qualitatively distinct sense of vivid, tangible, 3-D that it feels as though we can touch. This sense, referred to as stereopsis, is typically not present when viewing 2d paintings, or even photographs (Howard and Rogers, 2002). It is induced by viewing 3-D scenes defined solely by binocular disparity or motion parallax, however (Rogers and Graham, 1982). More recent work suggests that relative blur can also induce stereopsis (Vishwanath and Hibbard, 2013), again suggesting the contribution of focus cues may be greater than has typically been assumed.

### 1.5.3 Effect of incorrect presentation of focus cues in stereo 3-D

In recent years the quality that can be achieved in stereo 3d imagery has improved significantly due to improvements in areas such as the spatial resolution and dynamic range of display technologies, combined with continued developments in photorealistic rendering techniques (de Silva et al., 2011; Banks et al., 2016; Zhong et al., 2021). An exception is presenting focus cues correctly, which remains challenging. Conventional stereo 3-D displays present images at a single, fixed focal distance (the screen surface). Consequently, focus cues resulting from natural viewing, and from viewing an otherwise equivalent scene depicted in stereo 3-D, differ substantially. The stimulus to accommodation does not vary naturally with variations in scene depth but is instead consistent with the display surface. Moreover, the patterns of defocus blur that naturally result when scene points are nearer or farther than the currently focused distance are not reproduced correctly.

The finding that focus cues do affect perception of 3-D scene structure (see above) means that presenting focus incorrectly can cause distortions in perceived depth. Because focus cues are consistent with the screen surface, rather than the depicted scene, they would be expected to result in biases in perceived depth accordingly—i.e. a contraction in perceived depth around the screen—in accordance with the predictions of cue integration theory (see earlier). Incorrect focus cues also cause other issues, which we briefly outline below.

*Vergence–accommodation conflicts*

A comparatively well-studied aspect of conventional stereo 3-D presentation is the effect of vergence-accommodation conflicts. Humans achieve single, clear binocular vision by converging and focusing their accurately on points in the scene. In natural viewing the stimulus to vergence (disparity) and to accommodation (defocus blur) are consistent, and so require consistent responses. Indeed, the two processes are neurally coupled such that they drive one another synergistically (known as disparity-driven accommodation and blur-driven vergence) (Fincham & Walton, 1957; Martens & Ogle, 1959; Schor et al., 1992). When viewing a natural scene, the coupling helps to increase the speed of the response, such that both vergence and accommodation responses are faster binocularly than when viewed monocularly (Cumming & Judge, 1986; Krishnan et al., 1977; Semmlow & Wetzel, 1979). In stereo 3-D, however, the vergence stimulus varies with the depth in the depicted scene, whereas the accommodation stimulus is fixed at the screen surface. This vergence-accommodation conflict requires the observer to decouple the two responses in order to achieve clear, single binocular vision. This decoupling is effortful, and is not always possible, resulting in blurry and/or unfused binocular images (Wann & Mon-Williams, 2002; Howarth, 2011; Kooi & Toet, 2004; Lambooij et al., 2009; Banks et al., 2012; Urvoy et al., 2013).

Vergence-accommodation conflicts have been shown to have a number of unwanted effects in stereo 3-D viewing, which we briefly review below.

*Visual discomfort and visual fatigue associated with vergence-accommodation conflict*

To determine the effects of vergence-accommodation conflicts on visual comfort unambiguously it is necessary to compare stimuli presented with correct focus cues (as per the real world) and with conventional 3-D stereo viewing, where the stimulus to accommodation is fixed, while holding all other stimulus properties constant. Otherwise, any apparent adverse effects of stereo 3-D could be due to implementation issues such as crosstalk, increased sense of self-motion, decreased brightness, motion artifacts, or even the need to make vergence eye movements that is not present viewing 2-D images (Kooi & Toet, 2004; Palmisano, 1996, 2002; Hoffman et al., 2011). Hoffman et al. (2008) used a multiple-focal-planes display, using mirrors and beamsplitters (Akeley et al., 2004), which created a stack of transparent focal planes. This allowed the vergence and accommodation distance to be controlled independently to create conventional stereo 3-D and 'real-world' or correct-focus-cues conditions, while keeping other properties of the stimulus constant (Hoffman et al., 2008). They evaluated visual discomfort while following viewing of stereoscopic stimuli and carrying out a psychophysical task. Observers experienced higher visual discomfort and fatigue under the conventional stereo 3-D condition, demonstrating that vergence-accommodation conflicts per se cause visual discomfort (Hoffman et al., 2008).

Shibata et al. (2011) went on to quantify how discomfort was affected by screen distance, and the sign of vergence -accommodation conflicts (i.e. converging nearer vs farther than the screen). This study also used a multiple-focal planes display, but created the focal planes in a time-multiplexed fashion, using a high-speed switchable lens system (Love et al., 2009). Again, this allowed them to manipulate focal distance and vergence distances independently, as per Hoffman et al. (2008). Shibata et al. (2011) used their data to estimate a 'zone of comfort' for stereoscopic viewing, that described a range of depths (vergence distances) around the screen that could be comfortably viewed. This zone was found to be slightly larger at near distances compared to far. They also found that content beyond the screen was more comfortable at near distances, and content nearer than the screen was more comfortable at far distances. Stereoscopic displays with various viewing distances are commercially available and these data provide useful guidelines on how stereo content should best be configured for different viewing situations (e.g. handheld device vs. computer screen vs. cinema).

*Effects of vergence-accommodation conflict on stereo performance*

Vergence-accommodation conflicts have also been shown to degrade stereoscopic depth perception in various ways. They result in reduced stereoacuity (i.e. reduced ability to determine fine details in 3-D stereoscopic depth; Watt et al., 2005b; Hoffman et al., 2008). They also increase the time required for stereoscopic fusion to occur (Akeley et al., 2004; Hoffman et al., 2008). These results presumably reflect a combination of inaccurate or unstable vergence and/or accommodation caused by the requirement to decouple vergence and accommodation responses. Note that for an image to be sharply focused on the retina the eye must accommodate at the focal distance of the object within a margin of error of ±0.25-0.3 D (Campbell et al., 1957; Charman & Whitefoot, 1997), and for an object to be binocularly fused the eyes must converge at the correct vergence distance within a margin of error of 15-30 arc min (often referred to as Panum's fusion area; see also Blakemore, 1970; Ogle, 1932; Schor et al., 1984).

*Comparison to vergence-accommodation conflicts induced by optical prescriptions*

It is reasonable to query whether issues related to vergence-accommodation conflicts in stereo 3-D are unique to this setting, and require unique solutions, given that patients with new optical corrections are also exposed to vergence-accommodation conflicts, and can experience asthenopia symptoms and visual discomfort (Sheedy et al., 2003). Vergence-accommodation conflicts are introduced by optical correction changes to focal power because the focal demand on the crystalline lens is changed without changing vergence demand. Similarly, prescription of prisms can alter vergence demand. This creates a difference in the magnitude of the stimulus to vergence and accommodation, therefore creating a conflict. Patients eventually adapt to these changes, so might people become adapted to vergence-accommodation conflicts in stereo 3-D? A key difference is that the conflict introduced by optical corrections is constant, and so can be adapted to, whereas the conflict in stereo-3-D is typically constantly varying, and so adaptation may not be possible (Shibata et al., 2011).

*Age-related differences in individuals' tolerance to vergence-accommodation conflict.*

The decline in the eye's ability to adjust its focal power with age is attributed to the stiffening of the crystalline lens, a phenomenon termed presbyopia (Fisher, 1973; Glasser & Campbell, 1998; Mordi & Ciuffreda, 1998; Mordi & Ciuffreda, 2004). Divergent viewpoints exist regarding how the aging process itself and age-related changes in accommodation capability might influence the tolerance to vergence-accommodation conflict. There are conflicting ideas about how age per se, and age-related changes in the ability to accommodate, may affect tolerance to vergence-accommodation conflicts. One argument is that younger viewers will be more robust to the effort required to decouple vergence and accommodation responses, and so will be more tolerant than older people to vergence-accommodation conflicts (Mendiburu, 2009; Banks et al., 2012). Another argument predicts the opposite. A presbyopic person's real-world oculomotor responses—varying vergence but (near) fixed accommodation—closely match the requirements of S3D viewing. Therefore, provided that older adults use appropriate optical corrections for the screen's distance, they might actually manifest greater tolerance to vergence-accommodation conflicts than their younger counterparts (Banks et al., 2012). Supporting this latter notion, Yang et al. (2011) discovered that individuals aged 24-34 reported more pronounced adverse effects, as documented through a symptom questionnaire, compared to viewers aged >45 years when exposed to S3D content in a home-cinema environment.

*Could incorrect focus cues affect appearance or realism of stereo 3-D scenes?*

The above discussion considers essentially ergonomic, or human factors consequences of vergence-accommodation conflicts. It is also possible that incorrect focus cues affect more subjective aspects of stereo 3-D perception. Altering the normal pattern of accommodation and vergence responses may be perceptible, and feel unnatural, to people. Moreover, the unnatural patterns of retinal blur, and even diplopia, that may accompany these responses may also make scenes appear unusual, and therefore less realistic than they otherwise would. Finally, viewers may be sensitive to the cue conflicts between focus cues and other depth cues, which could again create an experience that is different to viewing natural scenes. We explore this possibility in detail Chapters 3 and 4.

## 1.6 Display: Getting focus cues correct

For focus cues to be presented correctly both the stimulus to accommodation and retinal blur gradient should be correct. Conventional stereoscopic 3-D displays do not get either of these aspects correct (Pastoor & Wopking, 1997; Sexton & Surman, 1999). Various approaches have been explored to achieving this, and which aim for varying degrees of correctness. Since all approaches to presenting correct focus cues require specialist technology, with associated costs, it is important to explore both the comparative benefits of different approaches.

*Gaze-contingent rendering*

One approach to addressing this is gaze-contingent displays, which uses an eye tracker to track viewer's gaze and perform a depth-dependent blurring of the region in the images that are nearer or farther than the currently fixated point (Duchowski et al., 2004; Biebl et al., 2022). This approach can be implemented with existing technology, but does not get the stimulus to accommodation correct as the image is still coming from the same, fixed display surface independent of the depicted depth. A further development of this idea is so-called varifocal gaze-contingent displays, which include a variable focus element, so that the display plane is moved to the optical distance of the currently fixated scene point (Hasnain et al., 2019). Such displays in principle stimulate accommodation near-correctly, However, a practical problem with all gaze-contingent rendering approaches is that they require very high precision and reliability of eye tracking to be able to detect small changes in vergence, and correctly identify the gaze position at scene points with large depth gradients, or in the case of transparency, where accommodation distances in very similar visual directions. It also remains to be determined how correct the fine details of blur rendering need to be to adequately approximate the real work. It has been shown that we are sensitive not only to first-order blur but also to chromatic aberration signals, for example (Nguyen, et al., 2005; Cholewiak et al. 2017).

There have been many attempts to create displays that can get focus cues more fully correct (Favolora et al., 2002; Lucente, 1997; McQuaide, 2002; Schowengerdt & Seibel, 2006; Sullivan, 2004; Akeley, et al., 2004; MacKenzie et al., 2010; Chang et al., 2018; Zhong et al., 2021). These are called volumetric displays and are classified into autostereoscopic and fixed-viewpoint volumetric displays (Blundell, 2012). Volumetric displays form a visual representation of a scene in three dimensions by placing light sources in a 3-D volume as compared to conventional displays that uses offset images that are displayed to each eye separately and brain combines then to produce a 3-D percept (Blundell, 2012). Volumetric displays provide focally accurate holographic wavefronts to the eyes via rotating display planes and stalks of switchable diffusers (Fanlora et al., 2002; Sullivan et al., 2004). This type of display can provide accurate representation of focus cues, stereo cues and parallax cues. These displays create a volume of pixels called voxels in 3-D space. However volumetric displays have several drawbacks such as small display volume, limited effective resolution, unable to produce occlusion effects, and other view dependent effects such as reflections.

### 1.6.1 Autostereoscopic display

There are basically three types of autostereoscopic displays in use: holographic, static-volume, and swept-volume displays. In all these displays the image points are presented within a 3-D volume, which provides matched cues to vergence and accommodation for the range of distances within the display (Akeley et al., 2004). A key advantage of these displays is multiple viewers can use the screen from multiple angles without glasses.

Holographic displays use computer-generated inference patterns of light to reproduce the properties of light waves (luminance, wavelength and phase differences) which allows generation of a 3-D scene (Pastoor and Wopking, 1997). Swept volume displays rapidly move the entire screen/display or uses a projection screen that spins on a central axis, and presents correct stimulus to accommodation. Images are built up in a time multiplexed manner by synchronizing display information with the position of the display element (Favalora et al., 2002; Schowengerdt & Seibel, 2006). In a static volume displays system multiple display planes are used to fill up a volume of space, which becomes the viewing space (Sullivan, 2004). Some displays under this category use liquid-crystal scattering [LCSS] (Sullivan, 2004), and out of the many displays only one is active a given time and acts like a scattering rear projection screen. The 3-D image is presented to each screen in the stack in a sweeping (a sweep through all the screens) manner creating a 3-D image in appropriate locations in physical space of the screen (Sullivan, 2004). However, most of these displays do not present stimulus to retinal blur corrects (esp. swept-volume displays). All these displays require high resolution in all directions, which is costly to compute. The range of vergence and accommodation distances that can be projected is limited and constraint to the volume of the display. These displays cannot correctly represent view-dependent changes in the scene such as occlusions, reflections. transparency because voxels emit the same light in all directions (Akeley et al., 2004).

### 1.6.2 Multiple-focal-planes displays

Multiple-focal-planes displays use one of several methods to create a situation where each eye sees the sum of images presented on image planes at different focal distances (see *Our display*, below). This allows focal distance to be varied by placing images on different focal planes. Most iterations have separate optical paths for each eye, known as a haploscope configuration. One method for creating multiple focal planes is to use beamsplitters and mirrors (as in *Our display*) to create a stack of continuously displayed, transparent image planes (Akeley et al., 2004; MacKenzie et al., 2010). An alternative approach is to present each focal plane in a time-multiplexed fashion, using a variable-power optical element synchronised to a high-speed display (Liu et al., 2008; Love et al., 2009; Shevlin, 2005) . Both approaches have constraints on the number of focal planes that can be displayed, and hence their resolution in focal depth. In beamsplitter-based systems the maximum luminance at a given image plane is given by the luminance of the monitor, divided by the number of image planes (MacKenzie et al., 2010) and so large numbers of image planes results in unacceptably low luminance. In time-multiplexed systems increasing the number of planes reduces the 'duty cycle' of each image plane, so results in similar reduction in effective luminance. And in general, increasing the number of focal planes increases computational overhead (Akeley et al., 2004). For these reasons, this approach requires a way to interpolate between image planes. Akeley et al. (2004) proposed that this can be achieved by distributing image intensity across image planes, for each point in an image, according to its dioptric distance from neighbouring image planes. This approach is referred to as depth filtering or depth-weighted blending. It has been shown that this approach can stimulate accommodation responses to intermediate distances between focal planes with plane separations up to at least 0.67 D (MacKenzie et al., 2010; 2012), making it a promising approach for addressing vergence-accommodation conflicts. The resulting retinal blur gradient differs from natural viewing in key ways, however, and multiple-focal-planes display in general are not able to present occlusion boundaries correctly (Ravikumar et al., 2008; Narain et al., 2015; Zannoli et al., 2016). There are optimised rendering approaches which can partially address these issues (Narain et al., 2015).

### 1.6.3 Light field display

Light field displays work on the principle of plenoptic function (Gershun et al., 1939, and emits four-dimensional distribution of light rays (Banks et al., 2016). These light rays vary across the two dimensions of the display, and over the horizontal and vertical viewing angles of each pixel. In a light field display stereo cues are motion parallax cues are accurate as both eyes receive different light ray when stationary and in motion, and because different parts of the viewer's pupil receive different light rays the focus cues are accurate too (Banks et al., 2016). However, these displays lack image resolution, as adding more light-field viewing zones reduces spatial display resolution. To create a high-quality light field at a refresh rate of 60 Hz, it would require rendering trillions of light rays, and the memory required to handle the data is immense and prohibitive (Banks et al., 2016). So, currently light field displays are not practically viable.

In fact, any of the stereoscopic displays that prove correct stimulus to focus cues is either commercially unviable mostly due to complex and expensive hardware, or they produce conflict free imagery on the expense of resolution and other view dependent properties in the image.

## 1.7 Our Display

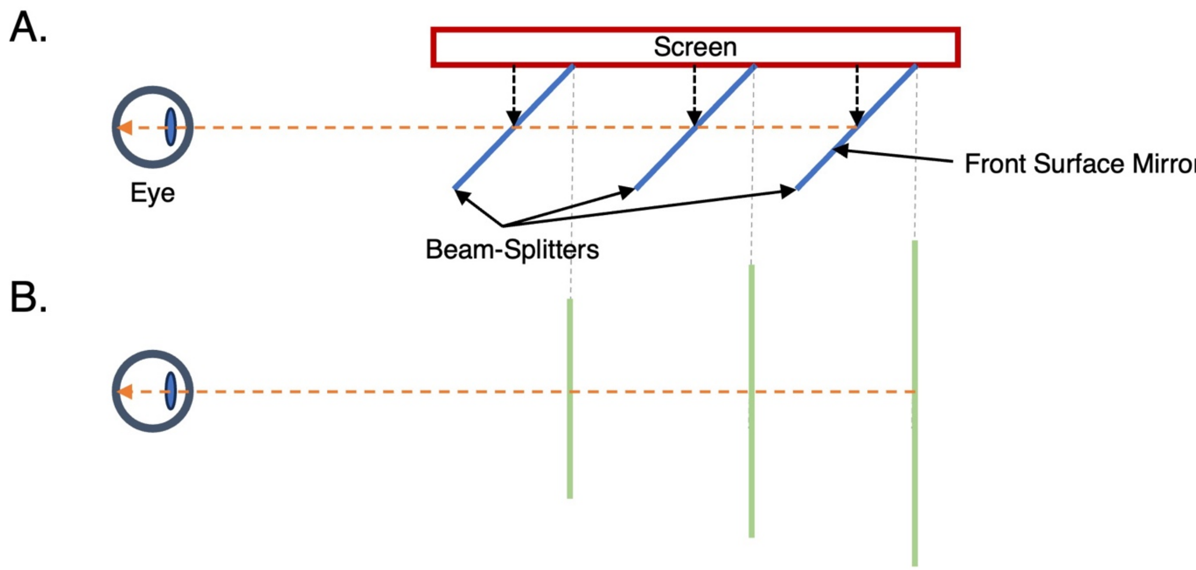



**Figure 1.8**: Side view of the Multiple-focal-planes display (one section). (A) Shows the monitor and the beam-splitters/mirror arrangement. The black dashed lines demonstrate the optical path for each focal plane. (B) Is an illustration of the resultant transparent focal planes; the grey dashed lines demonstrate the distance for each focal-plane in the display.

We used a three-plane multiple-focal-planes display for the experiments in Chapters 2 and 3 (MacKenzie et al., 2010) to manipulate the stimulus to accommodation in stereoscopic images, and present correct focus cues (in Chapter 3). Although the display was similar to that of Akeley et al. (2004) it differed in several key regards. We had moveable focal plane positions, and the use of a Badal optical system allowed various useful properties described below (Figure 1.8.A).

The images were displayed on 30” Samsung 305T TFT monitors which had a resolution of 2560 x 1600 pixel and 0.25 mm dot pitch. Each monitor was directly above one front-surface mirror and two beam splitters, so the images displayed on the monitors reflected from them, creating a stack of three transparent image planes at three focal distances (the mirrors were at an able of 45 degrees to the screen) (Figure 1.8.B).

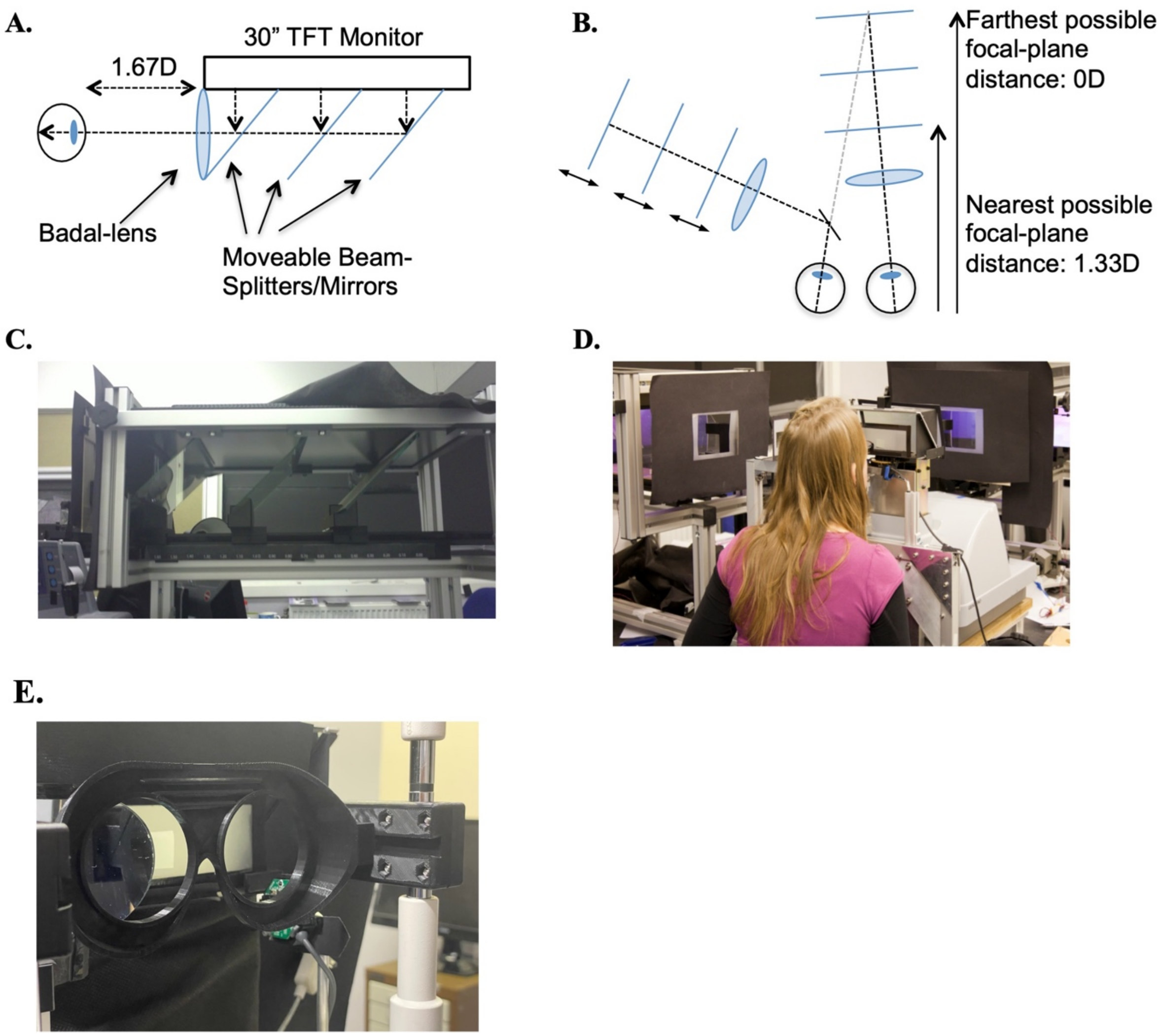



**Figure 1.9**: The multiple-plane display used in this thesis. (A) A side view of one eye’s display, showing the arrangement of beam-splitters, first-surface mirror and Badal lens. (B) Plan view of the arrangement of the left- and right-eye’s displays. The left eye viewed its display via a first-surface mirror; the right eye viewed the display directly. (C) The side view of the display showing the screen projecting onto the beam-splitters and first-surface mirror positioned on top of an optical rail. (D) An observer on the display. (E) The viewing point where observers viewed the display.

Since these mirrors and beam-splitters were mounted on optical rails, the separation between the focal planes was independently adjustable (Figure 1.9. A). Each eye had a separate monitor and focal-planes system (haploscope configuration) (Figure 1.9.B), making it possible to independently control the vergence and accommodation distances. The left eye views its display through a first-surface mirror near the eye, and the right eye views its display directly (Figure 1.9.B).

Each side of the display is mounted on a movable roller (ball) wheels, allowing horizontal movement when loosened. Thus, we were able to adjust the haploscope separation to match the inter-ocular distance (IOD) of the observers. This allowed us to create correct geometric locations of the left and right eye images on the focal planes, and position the eyes correctly relative to the Badal optical system (see below). Each eye's display can rotate around the eye's canter of rotation, allowing optimization of the binocular overlap for different screen distances.

This display allows us to independently control accommodation and vergence distance, allowing us to both simulate conventional stereo 3-D presentation and real world (correct focus cues) viewing) for certain configurations of stimuli, while holding all other stimulus properties constant.

### 1.7.1 Badal lens

Each eye's display had a Badal lens (positive spherical lens [1.67D]) through which the display is viewed (Figure 1.9.A). The positive spherical lens is placed at 60 cm (which is the focal length of the lens) from the eye. The lens was 15 x 12cm high, so it provides a maximum possible field-of-view of 14.3 x 11.4 degrees.

Placing the lens at its focal length from the eye creates a Badal optical system. This has several useful properties for our purposes. First, images have a constant angular size independent of focal distance. This simplifies stimulus calibration, and ensures that that display resolution is matched in angular terms at all focal distances. Second, in a Badal system, focal distance is linear in physical distance (unlike the normal case where distance in dioptres is the reciprocal of distance in metres). This makes setting and adjusting focal plane separations straightforward (especially helpful in our realism experiment in Chapter 3). Third, in a Badal system an image at the lens' focal length behind the lens is at optical infinity, allowing us to present stimuli at a wide range of distances within a physically small display (please refer to, MacKenzie, et al., 2010, for more details). All calculations for determining the geometrical image can be made with respect to a single virtual screen surface at the distance of the lens.

### 1.7.2 Viewing port

It is important to stabilize the participants' heads while viewing a multiple-focal-planes display, so that the image planes remain aligned. It is also important to know where the participants eyes are positioned to the display surface, so that stimulus position can be calculated and presented appropriately. We stabilized our participants using a forehead and chin rest, which incorporated a 3-D printed viewing port (see Figure 1.9.E).

### 1.7.3 Luminance calibration and alignment

Since we used multiple beam splitters for the purpose of creating a multiple-focal-planes system, we also introduced a variability in the luminance and wavelength characteristics of the monitor across focal planes. Because we needed to match stimulus properties across different focal planes, we carefully calibrated each eye's display for all image plane positions, to ensure luminance and colour were matched. To do this, we measured the overall 'gamma functions' (using a Minolta CS-100 Chroma Meter) by measuring the RGB triads required to produce a white square that fits in the display's field-of-view that had the CIE chromatic coordinate of R = 0.333, G = 0.333, and B = 0.333 at luminous intervals of 0.5-1 $cd/m^2$ at all focal planes. We then fit polynomial equations to each colour function allowing us to interpolate between measured luminance values to produce the target luminance required in each instance. Our overall maximum luminance level was determined by the focal plane with the lowest maximum luminance.

The head was aligned to the display by placing a sighting device in-front of the Badal lens which consisted of two thin wires that intersected at the centre of each displays' field of view. Participants centered a dot (10 arc min) to the nearest plane first (like a vernier task) and then to the consecutive planes by keeping the nearest plane as a reference for the mid plane, and mid plane as a reference for the far plane. This was done three times for each focal plane and the settings averaged. This alignment procedure was necessary for eliminating small but significant misalignments between the participants optical center and the display, and also to ensure that the focal planes were properly aligned with one another.

### 1.7.4 Software for running the display

The main computer CPU that controlled the image rendering and presentation ran on Macintosh. The image presentation and response collection on the display were done using MATLAB software (Mathworks Inc., Natick, Massachusetts) and a third-party psychophysics toolbox (Psych Toolbox, version 3.0.8, Brainard, 1997; Pelli, 1997).

## 1.8 Conclusion

The research on the effects of incorrect focus cues has mostly centered around vergence-accommodation conflict and associated visual fatigue. And there have been no significant work done to evaluate the effect of focus cues on perceived naturalness of depth in a scene (Depth realism). The work described in section 1.5 and 1.6 demonstrates that, there are observable issues with 3-D shape perception (Buckley & Frisby, 1993; Frisby et al., 1996; Frisby et al., 1995; Watt et al., 2005, Hoffman et al., 2008), visual fatigue (Ukai, 2007; Wann & Mon-Williams, 1997, Shibata et al., 2011), reduced stereoscopic perception (Watt et al., 2005; Hoffman et al., 2008), while viewing a stereoscopic 3-D display, implying that getting stereo and parallax cues correct, alone, does not guarantee a naturalistic (realistic) 3-D viewing experience. And it is highly likely that getting focus cues correct is of paramount importance for depth realism. So, in this project we have tried to evaluate if incorrect focus cues reduce the realism and naturalness of perceived depth for simple planar stimuli (Chapter 3).

And as a follow up work, we also evaluated if the lack of correct focus cues reduces depth realism for hi-fidelity images where focus cues might be barely detectable, and have practically very little contribution to depth realism (Chapter 4).

However, it is also pretty evident from section 1.6, that getting focus cues correct is hard (Akeley et al., 2004), as it involves developing complicated hardware and would be commercially unviable for the time being.  So, understanding how to mitigates the effects of incorrect focus cues (esp. vergence-accommodation conflict) in conventional stereoscopic 3-D displays would help us optimize stereoscopic 3-D display effectively.

A practical approach would be to determine the distances around the display, where people have acceptable tolerance to vergence-accommodation conflict. I would also be useful to test any factors such as Phoria (Shibata et al., 2011), ZCSBV (Shibata et al., 2011), and or age (Yang et al., 2018, have found interaction between display dimension and age on viewing symptoms and perceived immersion), that might predict individuals' tolerance to vergence-accommodation conflict. So, with all intents and purposes we decided to first evaluate, if any visual factor predicted individuals' tolerance to vergence-accommodation conflict induced degradation in stereo depth perception (Stereoacuity) in the next chapter (2).

# Chapter 2

## 2 Tolerance of stereoscopic depth perception to vergence-accommodation conflict in stereo displays is not predicted by age-related changes in the ability to accommodate

### 2.1 Introduction

The range of applications for stereoscopic 3-D (S3D) imagery continues to grow, and there is rapid development in the technology used to present it (including Virtual Reality and Augmented Reality, collectively known as Extended Reality or XR) (Lucente, 1997; Suyama et al., 2001; Favalora et al., 2002; McQuaide, 2002; Takaki, 2003; Sullivan, 2004; Akeley et al., 2004; Schowengerdt & Seibel, 2006). It remains the case that commercially available S3D display technologies typically introduce vergence-accommodation conflicts, however, by presenting images with natural variation in disparity-specified depths, but a single focal distance (the display surface; Watt et al., 2005; Hoffman et al., 2008; Lambooij et al., 2009). Vergence and accommodation responses are neurally coupled, and drive one another synergistically (Fincham & Walton, 1957; Martens & Ogle, 1959), and it is well established that the requirement to decouple them to view S3D imagery causes a range of adverse effects (Banks et al., 2012; Lambooij et al., 2009; Howarth, 2011). Perhaps the most well-recognised problems are viewer fatigue and discomfort (including eye strain, headaches etc.), thought to result primarily from the effort of trying to converge at one distance while accommodating to another (Inoue et al., 1997; Hoffman et al., 2008; Shibata et al., 2011). Because it is not always possible to decouple vergence and accommodation responses completely, one or both responses can also be inaccurate, causing double vision (diplopia) and/or blurred vision, resulting in degraded stereoscopic depth perception (Watt et al., 2005a, 2005b; Hoffman et al., 2008). Vergence accommodation conflicts lead to both decreased precision of stereoscopic depth perception (i.e., poorer stereoacuity; Watt et al., 2005; Hoffman et al., 2008), and slower stereoscopic fusion (Akeley et al.

2004; Watt et al., 2005b; Hoffman et al., 2008), compared to when vergence and accommodation stimuli are consistent (as per real-world viewing).

Although there are several potential technological solutions for minimising or even eliminating vergence-accommodation conflicts (e.g. Lucente, 1997; Suyama et al., 2001; Favalora et al., 2002; McQuaide, 2002; Takaki, 2003; Sullivan, 2004; Akeley et al., 2004; Schowengerdt & Seibel, 2006; Ravikumar et al., 2011; Huang et al., 2015; Johnson et al., 2016; Konrad et al., 2016; Kramida, 2016; Chang et al., 2018; Dunn et al., 2018), at present they remain experimental, and some barriers to their widespread adoption are likely to remain for the foreseeable future. Most approaches require specialist hardware (such as dynamic optics), resulting in increased costs and complexity compared to conventional displays (Rathivavel et al., 2018; Chang et al., 2018; Zhong et al., 2021). Moreover, many approaches currently achieve variations in focal distance at the cost of degrading other desirable image properties, for example reducing spatial resolution and/or dynamic range (Akeley et al., 2004; Lin et al., 2008; Lovel et al., 2009; Mackenzie et al., 2010). There are also everyday settings for S3D in which specialist viewing devices are impractical, such as cinema. Understanding how best to manage and mitigate effects of vergence-accommodation conflict in conventional displays therefore remains an important part of optimising the effectiveness of S3D imagery.

The relationship between magnitude of vergence-accommodation conflict and the severity of adverse effects is continuous rather than discrete: larger conflicts result in systematically worse effects (Watt et al., 2005; Hoffman et al., 2008; Shibata et al., 2010). A tractable approach to managing adverse effects is therefore to establish the range of stereoscopic depths around the screen plane (that is the range of vergence-accommodation conflicts) that result in acceptable levels of adverse effects. In other words, to determine the tolerance to vergence-accommodation conflicts. Specifically, determining the factors that predict this tolerance could allow the likely severity of adverse effects in S3D to be appropriately managed (Shibata et al., 2011).

One ‘class’ of predictor variable relates to the stimulus presented by S3D imagery—that is, properties of the content and display device. Shibata et al. (2011) measured how the screen’s focal distance, and the sign of the vergence-accommodation conflict (whether stereoscopic ‘objects’ were presented in-front-of or behind the screen) affected subjective ratings of discomfort. They found that the same conflict magnitude, in units of diopters, caused marginally more discomfort at a far screen distance (0.1 D or 10 m) compared to a near screen distance (2.5 D or 40 cm). They also found that the same dioptric conflict caused less discomfort behind the screen when the screen was near, and less discomfort in-front-of the screen when the screen was far. A further example of S3D stimulus properties predicting adverse effects is the finding that more rapid changes in vergence-accommodation conflict magnitude induce more discomfort (Kim et al., 2014). Such findings allow general practical guidelines for comfortable S3D imagery to be drawn up. For example, Shibata and colleagues derived a ‘zone of comfort’ for S3D imagery from their findings, which can be used to inform production of S3D content and development of displays (Shibata et al., 2011).

A second ‘class’ of potential predictors of tolerance to vergence-accommodation conflicts includes a number of visual-system characteristics that are implicated in viewing S3D imagery, and that vary across individuals. Such predictors offer the potential to tailor S3D content to specific users or groups, minimising adverse effects. Shibata et al. (2011) suggested that an individual’s *phoria* might be expected to influence where a person’s zone-of-comfort is positioned with respect to the screen. Phoria is the vergence posture adopted by the eyes when accommodated at a given distance, and when there is no effective stimulus to vergence (i.e. under monocular viewing, or when unfusable stimuli are presented to the two eyes). A person’s phoria need not correspond precisely to the accommodation distance, but can instead be nearer or farther (termed *esophoria* and *exophoria*, respectively). Phoria can be thought of as the resting state (i.e. zero-effort) posture of vergence for a particular accommodation distance. It might therefore be expected that people with esophoria at a given screen distance would be more comfortable with stereoscopically presented depths nearer than the screen rather than farther (and vice versa for people with exophoria). Consistent with this, Shibata et al. (2011) found a significant relationship between people's phoria and their zone of comfort. Shibata et al. (2011) also examined the relationship between the size of an individual's zone of comfort and their *zone of clear, single binocular vision (ZCSBV).* The ZCSBV is the maximum amount that vergence and accommodation stimuli can be placed in conflict (usually achieved using variable prisms) while still retaining binocular fusion and subjectively unblurred vision. As such, an individual's ZCSBV effectively represents their absolute ability to decouple vergence and accommodation responses, and so might be expected to predict the size of their zone of comfort. Again, Shibata et al. (2011) found evidence that this was the case (also see Shibata et al., 2011, for a review of how these measures relate to other measures of comfort developed for prescribing optical corrections).

Another visual-system characteristic that appears likely to play an important role in tolerance to vergence-accommodation conflicts is the ability to accommodate itself. The ability to change the focal power of the eye decreases as we age, due to stiffening of the crystalline lens—a condition referred to as *presbyopia* (Fisher, 1973; Glasser & Campbell, 1998; Mordi & Ciuffreda, 1998; Mordi & Ciuffreda, 2004). There are conflicting ideas about how age per se, and age-related changes in the ability to accommodate, may affect tolerance to vergence-accommodation conflicts. One argument is that younger viewers will be more robust to the effort required to decouple vergence and accommodation responses, and so will be more tolerant than older people to vergence-accommodation conflicts (Mendiburu, 2009; Banks et al., 2012). Another argument predicts the opposite. A presbyopic person's real-world oculomotor responses—varying vergence but (near) fixed accommodation—closely match the requirements of S3D viewing. So provided they are wearing an optical correction appropriate for the screen distance, older adults may in fact be more tolerant to vergence-accommodation conflicts than younger people (Banks et al., 2012). Consistent with this latter idea, Yang et al. (2011) found that people aged 24-

34 reported greater levels of adverse effects, reported via a symptom questionnaire, than viewers aged >45 years when viewing S3D content in a home-cinema setting.

Predicting tolerance to vergence-accommodation conflicts from age is potentially of more practical use than predictions based on phoria or ZCSBV. Age is a readily 'observable' variable, whereas few people know their phoria or ZCSBV (and these measurements are typically not particularly strongly related to other 'observable' variables, including age; (Srinivasan et al., 2015)). The expected age of a target audience could be used to tailor the range of vergence-accommodation conflicts presented, for example, so as to optimise tolerance. And rendered content (in VR etc.) could even be adjusted on the basis of each user's age. Note, while age and presbyopia are highly correlated, there is still substantial variation in the ability to accommodate at a given age, particularly below 50 years old. It is therefore important to understand not only how tolerance to vergence-accommodation conflict age is related to age, but also how it is related to the potential underlying variable—the ability to accommodate.

In the current study we examine how tolerance of stereoscopic depth perception to vergence-accommodation conflicts is affected by age, and the ability to accommodate. We measured tolerance in adults aged 18-75 years by determining how much conflict was required to result in a criterion-level reduction in stereoacuity. We measured this tolerance at several viewing distances, and for stereo stimuli in-front-of and behind the screen. We measured stereo depth perception performance, specifically, for several reasons. First, we wished to map out a 'zone of good stereo', analogous to Shibata et al.'s (2011) zone of comfort, because this has not previously been measured, and is an important functional aspect of effective S3D imagery. Second, measurements of discomfort and fatigue require comparative judgements made across long testing sessions with a fixed conflict (Shibata et al., 2011), and so are less tractable experimentally, particularly given that we wished to recruit from a wider-than-typical demographic. Our approach allowed us to estimated tolerance to vergence-accommodation conflict for an important aspect of stereo 3d perception within a relatively short testing session. Finally, the use of psychophysical methods allows precise, objective determination of an aspect of tolerance to vergence-accommodation conflict with minimal measurement noise, which should maximise the power to detect any relationships with other variables. Note, we do not view the 'zone of good stereo' as a proxy for a zone of comfort (further study is required to understand the nature of any such relationship between these aspects). Rather, measuring stereo depth perception provides a direct, objective measure of a pertinent aspect of responses to S3D imagery. We measured observers' ability to accommodate by measuring their accommodation stimulus-response function, following common clinical practice. This allowed us to examine the relationship between tolerance to vergence-accommodation conflict and not only age, but also the ability to accommodate per se. We also measured observers' phoria, and ZCSBV, using methods derived from standard clinical practice.

## 2.2 Method

### Observers

Fifty one observers took part in the study, aged 18 to 68 years (refer to table 2.1). Full details of observers' ages are reported in Results. All participants had stereoacuity in the normal range, as assessed by the Randot stereo test (Stereo Optical Company, Inc.), and normal or corrected-to-normal vision, as assessed by the Freiburg Visual Acuity & Contrast Test (FrACT; Bach,2007). The observers wore their optical corrections during the experiment, including during the measures of predictive factors (two observers wore contact lenses, 17 wore glasses with a single optical power, four wore varifocals; 28 observers had no correction). Observers gave informed consent and received financial compensation for their participation. All the observers were naïve to the purpose of the experiment. The experiment was approved by the departmental ethics committee of Bangor University and conducted in accordance with the Declaration of Helsinki.

| Gender | N | Mean Age (years) |
|---|---|---|
| Male | 23 | $34.43 \pm 14.69$ |
| Female | 27 | $41.37 \pm 15.14$ |
| Unspecified | 1 | 30 |
| **Total** | 51 | |

**Table 2.1**: Mean Age by Gender. Descriptive statistics for age by gender. Values are presented as means with standard deviations.

**Multiple-focal-planes display**

In order to vary focal distance and vergence distance independently we used a multiple-focal-planes stereoscopic display developed by MacKenzie and colleagues (MacKenzie et al., 2010, 2012) (refer to Chapter 1, section 1.7).

**Stereoacuity stimulus**

We measured stereoacuity under various experiment conditions (see below) using a random-dot stereogram stimulus, depicting a sinusoidal corrugation in depth (Figure 2.1). The 'bars' of the corrugations were oriented either plus or minus 15 deg from horizontal (i.e. the direction of disparity modulation was +/-15 deg from vertical) and participants judged whether the stimulus was oriented left-side-down or right-side-down (Figure 2.1). We varied peak-to-trough disparity of the corrugations to determine a threshold for when the 3-D structure was visible. The corrugation frequency was 1 cycle per degree, and the stimulus was clipped by a 4 degree diameter circular aperture (Figure 2.1). The dot density was 45 dots/deg$^2$ and each dot was 2 pixels wide (~3 arc min). The luminance of the dots was 25 cd/m$^2$ on a 'black' background 0.21 cd/m$^2$.


**Figure 2.1**: The stimuli were composed of random dot and not lines (as shown here for clarity) defined sinusoidal corrugation. The perceived corrugations were orientated either -15 degrees or +15 degrees from horizontal.

**Procedure**

*Overview*

To characterise each observer's zone of good stereo we measured the tolerance of stereo performance to vergence-accommodation conflict at each of the three focal distances. We reasoned long testing sessions and/or grossly exceeding an individual observer's tolerance risked causing 'after-effects' that could artificially reduce stereo performance even at subsequent small conflicts. Therefore, rather than measuring stereo performance at a series of (potentially too large) pre-determined conflicts we instead used an adaptive method, 'walking' the stimulus in-front-of and behind the screen until a criterion level of performance decrement was observed relative to when there was no vergence-accommodation conflict. So that this was comparable across observers, we first determined the peak-to-trough disparity that, with no vergence-accommodation conflict, resulted in each observer achieving a baseline stereo performance level of 90% correct. We then fixed the stimulus depth at this individualized value, and measured the magnitude of vergence-accommodation conflict required for their performance to drop to 75% correct. Thus, for each observer, we measured the magnitude of conflict that resulted in the same criterion change in stereo performance—from 90 to 75% correct—holding all other stimulus properties equal.

Observers completed the initial baseline test, then the zone-of-good-stereo measures, and finally the other clinically derived measures (see below) during multiple visits, spread across 2-3 days. This order avoided measurement of the ZCSBV, in particular, causing fatigue that could influence the zone-of-good-stereo-measures. Observers with corrections followed their normal pattern of use for different viewing distances. For example, if they normally wear their correction for screens at our farthest distance, but not for middle and near distances, they did the same in the zone-of-good-stereo test.

*Initial baseline with no vergence-accommodation conflict*

To measure baseline stereoacuity performance we measured corrugation orientation discrimination as a function of peak-to-trough disparity with no vergence-accommodation conflict. We did this separately for 1.3, 0.7 and 0.1 D focal distances. We used staircase procedures to vary the disparity level, and subsequently fitted a psychometric function to all of the data (a cumulative Gaussian, using a maximum-likelihood criterion; Wichmann and Hill, 2001; Kingdom, 2009). The 90% correct point was then derived from the fitted function (Figure 2.2). At each focal distance, observers completed three interleaved staircases in a single block, with different reversal rules: one repetition of a 1-down, 2-up staircase, and two repetitions of a 1-down, 3-up staircase, to distribute data points along the psychometric function (staircase chosen at random on each trial). The initial peak-to-trough disparity was 50 arc sec. The initial disparity step size was 5 arc sec. This was halved after the first four reversals, then halved again after four more reversals. The staircases terminated after 12 reversals.

Each trial consisted of an audible beep, followed by a fixation target (Maltese cross), presented for a random interval between 1 and 1.5 seconds. The stimulus was then presented for 2.5 seconds after which the observers made their response (oriented left-side-down or right-side-down?) by pressing a gamepad button. Responding initiated the next trial. Figure 2.2 shows an example of a fitted psychometric function and derived 90% correct disparity level. The data were analysed immediately and used to parameterise the stimulus for the zone-of-good-stereo measurement.

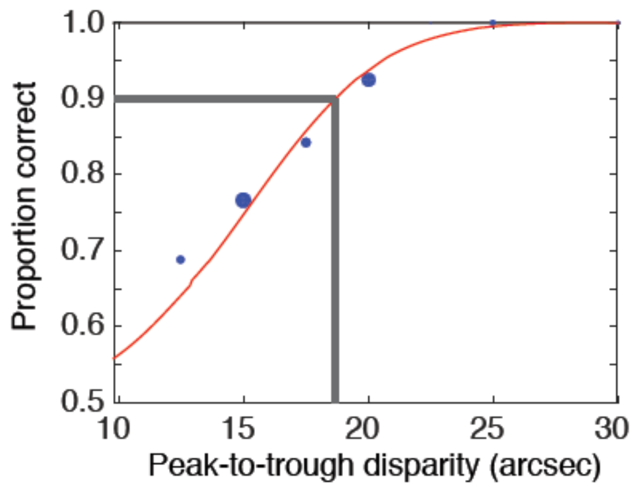



**Figure 2.2**: Example of a typical good fitted psychometric functions. The figures plot proportion correct at identifying the corrugation orientation, as a function of its peak-to-trough disparity. The size of the data points (blue circles) is proportional to the number of times the observer responded to that disparity. The red line is the best fitting cumulative Gaussian through the data. Dark grey line represents the disparity at which the orientation judgement was 90% of the times correct.

*Zone of good stereo performance*

We next measured each observer's zone-of-good-stereo. Here the peak-to-trough disparity was fixed at each individual's 90% correct level, however, and the magnitude of vergence-accommodation conflict was varied. Trials were blocked by focal distance. We presented conflicts in-front-of and behind the screen in the same block, chosen at random on each trial. The conflict magnitude was controlled by two separate, interleaved staircases—one for each sign of conflict (Figure 2.3). We interleaved conflicts in-front-of and behind the screen for two reasons. First, vergence-accommodation cross-links may adapt to a constant sign of conflict (Miles et al., 1987; Wann et al., 1995), altering the effects of a given conflict. Second, this condition better represents actual stereoscopic content viewing, where so-called positive and negative conflicts are both typically used. The staircases used a 1-up, 2-down reversal rule, where ‘up’ means the stimulus moved nearer to the screen, making the task easier (i.e. conflict was reduced), irrespective of the sign of the conflict. The ‘nearer’ and ‘farther’ staircases both started at zero conflict (Figure 2.3). The initial staircase step size was 0.1 D, halved after four reversals, and the staircases terminated after 12 reversals. The staircases were constrained to not ‘cross over’ to the other sign of conflict (that is, the in-front-of-screen staircase could not present stimuli beyond the screen). These staircases should converge around the 75% correct point, and so effectively ‘walk’ the stimulus in-front-of and behind the screen until the criterion-level drop in stereoscopic performance is reached. Observers completed three repetitions of each staircase, at each focal distance. Focal distances were randomly ordered. The observer’s task and stimulus timings were the same as for the baseline measure, above. The fixation cross always appeared at the focal (screen) distance, that is with no vergence-accommodation conflict.

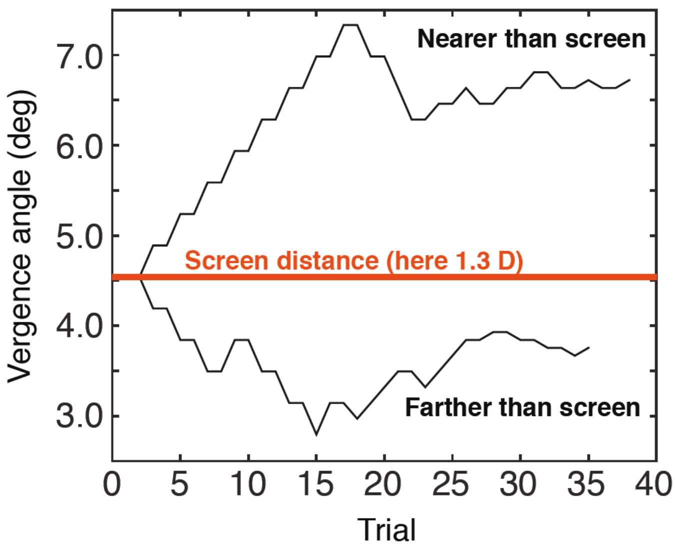



**Figure 2.3:** Example of staircase used for conflict manipulation. The figure plots conflict in vergence angle (degrees) across a block of trial. Two separate interleaved staircases present conflict in 1-up, 2-down (reversal rule) fashion, one for each sign of conflict (black line). The ‘nearer’ and ‘farther’ staircases both starts at zero conflict, on the screen (red). The staircases terminate after 12 reversals.

**Other visual-system characteristics**

*Ability to accommodate (degree of presbyopia)*

To determine observers' ability to accommodate we measured their accommodation stimulus-response function (variations in the amplitude of the accommodation response with changes in the focal distance to a stimulus). This function is typically linear with variations in stimulus distance within the range where people are able to vary accommodation. We characterised each observer's degree of presbyopia as the slope of this linear portion of their accommodation stimulus-response function (see Results).

We made subjective measurements of accommodation amplitude using a specially constructed stigmascope (Figure 2.4; Ciuffreda, 1991). The observer's right eye views a reference target, while simultaneously viewing a stigma (a point-light source), superimposed via a beam splitter (see Figure 2.4). The measurement works on the principle that the reference target is a good accommodation stimulus whereas the stigma is a poor accommodation stimulus, and so the accommodation response is driven by the target. The observer attempts to maintain accommodation at the target, while adjusting the stigma until it appears maximally sharp (i.e. optically conjugate to the observers' retina). This position corresponds to the current amplitude of accommodation (Ciuffreda, 1991).

Both target and stigma are viewed through separate Badal lenses (here 10 D, at 10 cm from the eye) (Figure 2.4). This prevents size changes or looming cues to distance resulting from movements in the target and stigma, and allowed presentation of focal distances from 7 D (14.3 cm) to 0 D (infinity) in a physically small device. The reference target comprised high-contrast letters (white letters on a black background), illuminated from above with a bright, diffuse light source (Figure 2.4). The stigma was created by a light box containing a white filament bulb, and with a 0.5 mm hole drilled in the front to create a point-light source (Figure 2.4). The exterior of the components was painted matt black. Both the stigma and the reference target were mounted on optical rails, to allow their focal distance to me moved towards and away from the observer's eye. A chin and forehead rest, adjusted for each observer, positioned the eye relative to the Badal lens system. The stigma was adjusted horizontally and vertically to superimpose it on a black portion of the reference target.

To measure the accommodation response the experimenter first adjusted the reference target to a given focal distance. The observer was then instructed to keep the letters clear and in focus while moving the stigma to the point where it looked sharpest, and the result was recorded manually. The accommodation response was measured for target distances from 0 to 7 D in 0.5 D increments. Each distance was tested twice. Observers were first tested at each distance in sequential order, from farthest to nearest (0 D to 7 D). They were then tested a second time, in reversed order (nearest to farthest; 7 D to 0 D). Prior to each measurement, the stigma was positioned a random distance from the reference target, but was positioned nearer/farther than the target an equal number of times. This was done to eliminate bias due to direction of movement of the stigma: because the eye has a non-zero depth of focus, the stigma can appear sharp at different points for inward and outward movements. We averaged across measurements taken with inward and outward stigma movements to give a single accommodation amplitude at each stimulus distance.

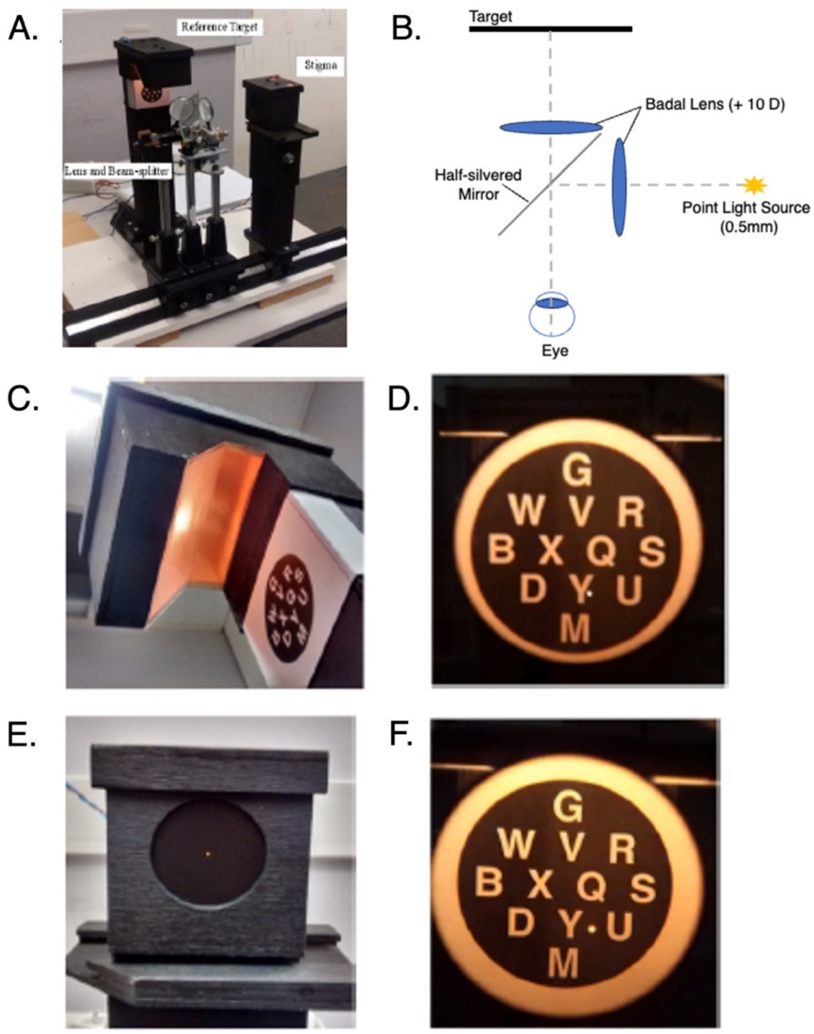



**Figure 2.4**: Images of the stigmascope used in this experiment. Image (A) is a photo of the device while (B) is an illustration, displaying the components of the stigmascope; this arrangement includes a beam-slitter and two convex lenses (10D). Image C is the stigma, light illuminating from above through a diffusing material to ensure an even distribution of light. Image E is the point light source (stigma), a pinhole is cut into the otherwise sealed box allowing the light to be directed along the line of sight. Images D and F show what the observer saw during the visual test, image D shows an unfocused stigma while image F shows a more focused stigma.

*Zone of clear, single binocular vision (ZCSBV):*

The ZCSBV was measured using a procedure based on clinical assessment (and similar to that used by Shibata et al., 2011). Observers wore goggles that placed Risley prisms (continuously adjustable laterally displacing prisms) in front of each eye, adjusted to match their IOD. Equal-and-opposite adjustments to the Risley prisms change the vergence stimulus while holding focal distance constant. To measure the ZCSBV observers viewed a column of high-contrast letters printed on paper and the vergence stimulus was continually adjusted in one direction (by the experimenter, at ~2 prism dioptres per second) until the observer reported that the letters appeared blurry and/or they experienced diplopia (double vision). Inward and outward measures were taken separately and averaged across. We measured the ZCSBV at the same three focal distances used in the main experiment (1.3, 0.7 and 0.1 D) using a long corridor. The letter stimuli were matched for angular size at the different viewing distances.

*Phoria*

We measured observers' horizontal phoria using a method based on the principle of the Maddox Wing, a handheld device used in clinical settings. This device presents different, unfusable images to the two eyes, so that there is no retinal stimulus to vergence, but there is a stimulus to accommodation. Phoria is the vergence posture adopted under these conditions. In the Maddox Wing, one eye sees a vertical arrow or pointer, and the other eye sees a horizontal measurement scale. To the observer, the arrow appears aligned with different points on the scale depending on their vergence posture. The observer is asked to report where they see the arrow, and from this the vergence posture can be calculated.

We created a virtual version of the Maddox Wing using our multiple-focal-planes display, allowing us to measure phoria at the same three focal distances used in the main experiment (1.3, 0.7 and 0.1 D). A vertical arrow was presented to the left eye, and a numerical scale was presented to the right eye (Figure 2.5). The arrow and scale were centred on their respective displays, and the display arms were rotated so that, at each focal distance, the lines-of-sight through the centre of each eye's display intersected at the tested focal distance. Therefore, if the person was converged at the focal distance the arrow would align perfectly with zero on the scale, indicating zero phoria. The numbers were spaced 1 degree apart. Positive and negative numbers were different colours, so that the sign of the observers' responses could be determined unambiguously. Observers were instructed to report the number the arrow was pointing to and the colour of that number (for example, "red 5"). They were told that the arrow may not be completely stationary, in which case they should report the values it moved between. The experimenter noted these and took the mid-point as the phoria value. Phoria was measured three times per focal distance, and the average taken.

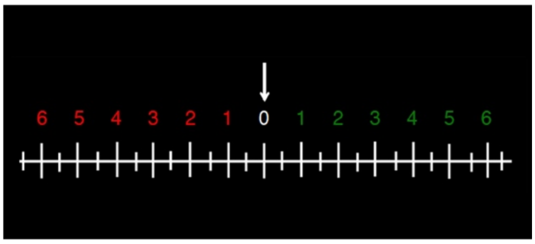



**Figure 2.5**: Illustration of virtual Maddox wing as viewed by participant during measurement of phoria. Right eye was presented with the scale and left eye presented with the arrow. The numbers are spaced 1 degree apart. The positive and negative numbers were different colours, so that the sign of the observer's response could be determined (e.g., "green 4").

## 2.3 Results

*Ability to accommodate*

As described above, we characterised each observer's ability to accommodate by determining the slope of the linear portion of their accommodation stimulus-response function. Figure 2.6 shows three examples of observers with varying ability to accommodate (i.e. varying degrees of presbyopia). The left panel shows a 20-year-old observer whose accommodation stimulus-response function has a slope of near 1.0, indicating the range of their accommodation response essentially matched variations in stimulus distance. The middle panel shows a 42-year-old observer whose response was linear until reaching a 'near point' at ~4 D, beyond which their accommodation did not change. The linearly varying portion of their response function had a shallower slope than the observer in the left panel, indicating a smaller range of accommodation for the same change in accommodation stimulus. The right panel shows a 67-year-old observer whose accommodation hardly changed with variations in accommodation stimulus, indicating presbyopia. The intercept value in this context serves as an indicator of the participants' ability to focus at optical infinity. The three observer's data plotted here happen to show correct focus at infinity, but this value typically varies based on the degree of refractive error exhibited by each participant (indicating individual differences in degree of myopia).

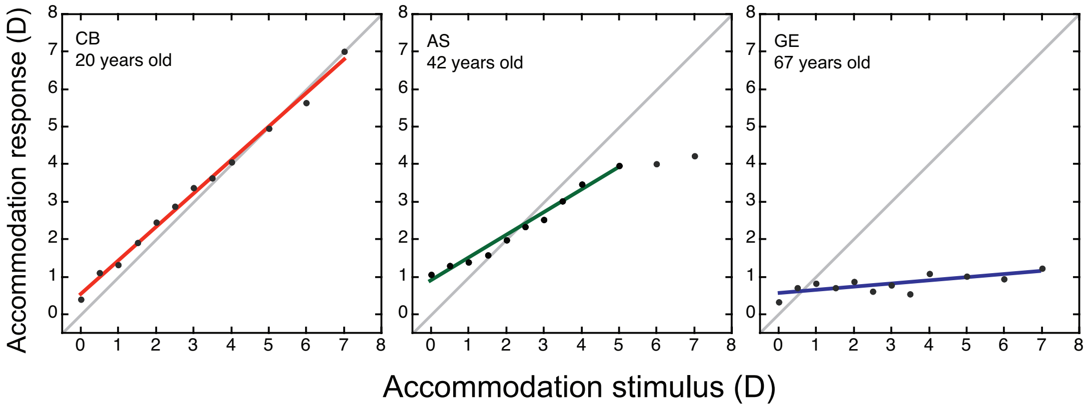



**Figure 2.6.** Accommodation stimulus-response functions for three example observers. Average accommodation response is plotted as a function of the accommodation stimulus. The left panel shows a typical young, non-presbyopic observer. The middle panel shows a middle-aged observer, with intermediate ability to accommodate. The right panel shows a typical older, presbyopic observer. The grey diagonal lines represent perfect accommodation response with respect to changes in stimulus distance. The coloured solid lines denote the best-fitting linear regression to the linear portion of the accommodation-stimulus response function in each case.

Figure 2.7, shows each observer's degree of presbyopia (the slope of their accommodation stimulus-response function) as a function of their age, for all observers except three who did not complete the accommodation stimulus-response measurement. Our data show the typical strong reduction in ability to accommodate with increasing age. A simple linear regression showed that this relationship was highly statistically significant ($R^2 = 0.73$, $F(1,46) = 128.03$, $p < 0.001$). The data for the youngest and oldest observers are tightly bunched around large and small slopes, respectively. There is a high degree of variability in ability to accommodate in observers aged ~30-50 years, however. This presumably reflects individual differences in the age of onset of presbyopia, and highlights how individual viewers in this age range age may be a poor predictor of the ability to accommodate.

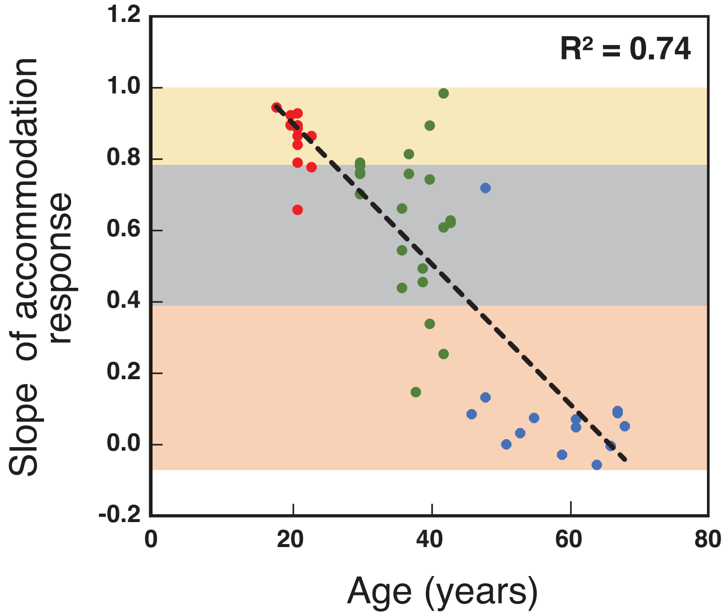



**Figure 2.7**: Relationship between ability to accommodate and age. Accommodation response is plotted as function of age. Participant were divided into three groups as per their age into (younger = red, middle-age = green, and older = blue). Participant were also divided into three groups based on their slope of accommodation response into (non-presbyopes = yellow, intermediates = dark grey, and presbyope = orange). The black dashed is the linear fit ($R^2 = 0.74$).

*Relationship between stereoacuity and age*

We initially determined a baseline for stereoacuity performance at each screen distance by determining the peak-to-trough disparity at which observers could identify the corrugation orientation at 90% correct (see Method). This was intended to establish a criterion for performance (and for reduction in performance to 75% correct with vergence-accommodation conflict) that was comparable across individuals. Nonetheless, it is useful to examine whether stereoacuity varied dramatically with age, in which case the comparison may be questionable. Moreover, it is informative to compare our data to previously observed measurements of stereoacuity variations with age (Zaroff et al., 2003; Haegerstrom-Portnoy, 2005).

Figure 2.8, plots stereoacuity as a function of age for each observer, averaged across the three screen distances. It can be seen that stereoacuity gets worse (larger depth required to discriminate the corrugation orientation) with increasing age. This is consistent with previous research, which showed 63 -75% of individuals in the age range 60-80 years have reduced stereoacuity**,** however, stereoacuity hugely varies between individuals in the age range of 60-80 (Zaroff et al., 2003), we also saw a similar trend in our stereoacuity data as well (see figure 2.8). The results of a linear regression analyses showed that while the relationship between age and stereoacuity was statistically significant, age accounts for only a small amount of variance in stereoacuity ($R^2 = 0.15$, ($F (1,49) = 8.4758$, $p = 0.005$).

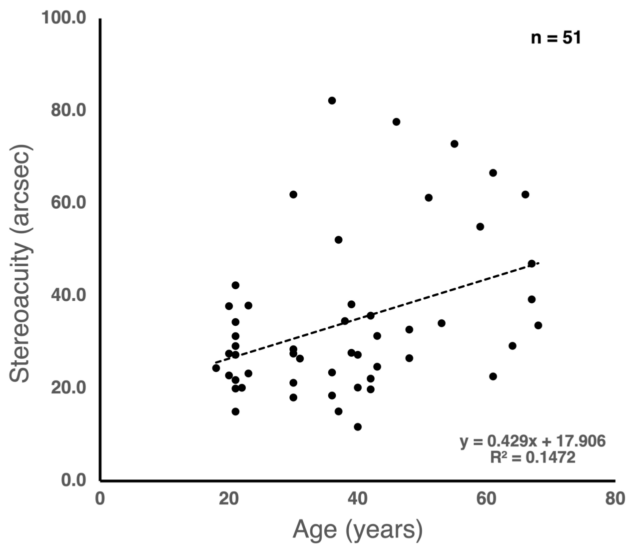



**Figure 2.8**: Relationship between stereoacuity and age. Stereoacuity (90% correct values) averaged across the three screen distances is plotted as a function of age. Pearson's R = 0.15

*Relationship between age and zone of good stereo perception*

In order to analyse whether age affected the properties of the zone of good stereo we first categorized participants into three groups according to age (see Figure: 2.7). We did this by rank ordering all participants by age, and using the location of the largest steps in age at appropriate points as the 'breakpoints'. This resulted in a younger group (n = 15; aged 18-23 years, mean = 20.93 years), a middle-age group (n = 22; aged 30-43 years; mean = 36.86 years), and an older group (n = 14; aged 46-68 years; mean = 58.14 years). The effects of age (and ability to accommodate) could be analysed either by treating the independent variable as continuous, or by categorising into groups. We chose to analyse the data broken into groups for pragmatic reasons. We did not know, a priori, what shape of function (i.e. linear regression, or a more complex function) would best fit the continuous data, and this would be difficult to determine from our data. Instead, the comparison of young vs. old groups speaks directly to our question and is straightforward, as it does not depend on assumptions about the continuous relationship between these variables.

We analysed the data separately according to the overall width of the zone of good stereo (i.e. the space encompassed by the inward and outward zone measurements), and the position of the zone with respect to the screen distance (Shibata et al., 2011). Figures 2.9 (A-C), show individual observer's inward and outward zone measurements, for each screen distance, as a function of age. The data are summarised in Figure 2.9 (D), which plots the overall width of the zone of good stereo, averaged within each age group, and screen distance. It can be seen that there are no clear effects of either age, or screen distance on the overall width of the zone of good stereo. To test the two alternative predictions statistically (that either older or younger participants would have larger zones of good stereo) we carried out an independent-samples t-test comparing the overall mean zone width, averaged across screen distance, for younger and older age groups. The result was non-significant ($t(27) = 0.606$, $p = 0.549$) confirming that there was no reliable effect of age on overall zone width. The effect size Cohen's d of 0.225 suggested a small magnitude of difference. The calculated statistical power ($1 - \beta$ error probability) was approximately 0.089. This suggests that our study had a relatively low statistical power to reliably detect the observed effect size (Cohen's d = 0.225) with the given sample size. This outcome underscores that our sample size might not have been sufficiently large to reliably detect correlations or differences of this magnitude.

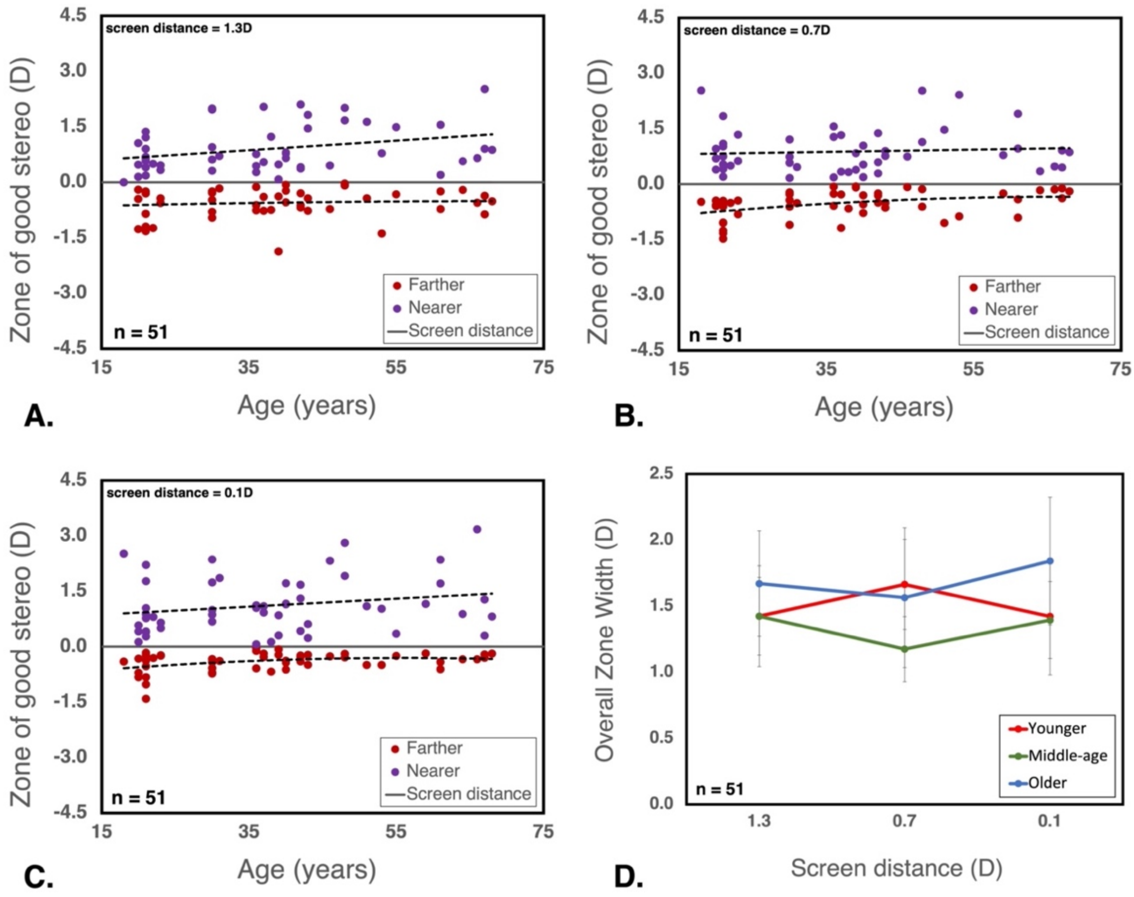



**Figure 2.9**: Relationship between zone of good stereo depth perception and age. Figures in panel A, B, and C, plots individual observer's zone of good stereo as a function of their age at three screen distances 1.3D, 0.7D, and 0.1D respectively. Individual datapoints in violet represent zone nearer to the screen (Positive zone) and data points in maroon represents zone farther from the screen (Negative zone). The dark black line indicates screen distance. The curves (dashed lines) are second-order polynomial fits to the data for the near and far zones. **D.** The overall zone width averaged across all observers is plotted as a function of screen distance for three age group. The younger, middle-age, and older are groups are indicated by red, green, and blur lines. The error bars are ± 1 standard error of the mean.

We also analysed the position of the zone of good stereo with respect to the screen surface, in order to determine whether it was centred or asymmetrically distributed around the screen (as Shibata et al., 2011, found for their zone of comfort). To do this, we computed the centre point of each observer's inward and outward measurements at each screen distance. The screen centre positions, averaged across all observers, are plotted in Figure 2.10. The data are plotted relative to the appropriate screen distance, so positive values indicate the centre of the zone of good stereo was nearer than the screen (and vice versa). Some clear patterns are evident in the plot. First, the zone centres were nearer than the screen at all three distances. Second, the plot shows evidence of effects of both screen distance and age group. In all age groups the zone of good stereo was increasingly shifted increasingly near with farther screen distances (indicating increasing asymmetry of the zone of good stereo with respect to the screen). This is consistent with the pattern observed by Shibata et al. (2011) for the zone of comfort. Moreover, the zone appears to be nearest for the older age group. Because we did not have specific predictions for this measure, we analysed the data by conducting a two-way (screen distance x age group) mixed-measures ANOVA. There was a significant main effect of both screen distance ($F(2,94) = 9.97$, $p < 0.001$) and age group ($F(2,47) = 7.41$, $p = 0.002$). The effect size was minimal with a partial eta squared value of 0.175 and 0.240 for screen distance and age group respectively. The screen distance x age group interaction was not significant ($F(4,94) = 0.49$, $p = 0.744$) with a minimal effect size (partial eta squared value of 0.020). Post hoc Bonferroni-corrected pairwise comparisons for the effect of distance showed that the zone centre was significantly nearer at 0.1 D than for both the 1.3 and 0.7 D distances ($p < 0.001$). The zone positions for 1.3 and 0.7 D distances were not significantly different ($p = 1.0$). Similar tests of the effect of age group showed that the older group's zone was significantly nearer than for both the younger ($p = 0.001$) and middle ($p = 0.021$) age groups.

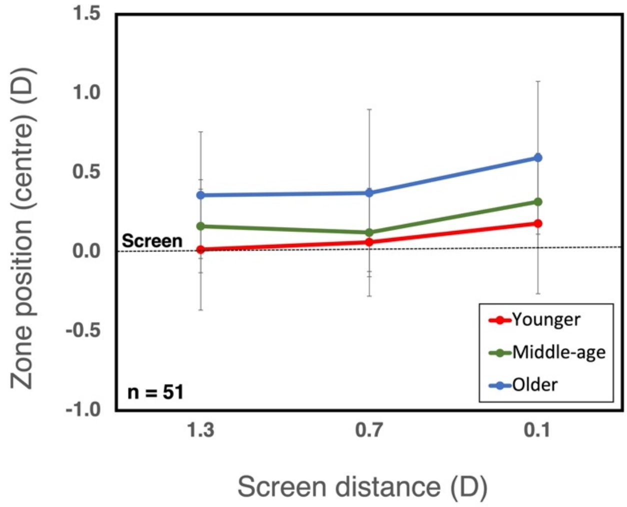



**Figure 2.10**: Relationship between position of zone of good stereo depth perception and screen distance for age groups. Zone centre positions averages across all observers is plotted as a function of screen distance from three age groups. The thin dashed line represents the screen. The younger, middle-age, and older are groups are indicated by red, green, and blur lines. The error bars are ± 1 standard error of the mean.

*Relationship between ability to accommodate and zone of good stereo perception*

We next repeated the above analysis, but as a function of observers' ability to accommodate, instead of age (see Figure 2.7 (B)). The slopes of the observers' accommodation stimulus-response functions were reasonably continuously distributed (without clear 'breakpoints'). We therefore categorised them by creating three equal-sized bins of 16 observers (note we had accommodation response data for 48 of 51 observers): the *non-presbyope* group had accommodation stimulus-response slopes from 0.98 to 0.779; the *intermediate* group had slopes from 0.776 to 0.438; the *presbyope* group had slopes ranging from 0.337 to -0.058. As Figure 2.7 shows the age and ability-to-accommodate groups had some overlap in membership, but were by no not identical. Most notably, the non-presbyope group included five observers from the middle age group.

Figure 2.11 plots the zone of good stereo in the same form as Figure 2.9, but as a function of ability to accommodate instead of age. Similar to above, no clear effects of ability to accommodate on zone width are evident in either the individual scatterplots, or when analysed by group (panel D). Consistent with this, the difference between overall average zone width for the non-presbyope and presbyope conditions (collapsed across screen distance) was not statistically significant ($t(30) = 0.115$, $p = 0.909$) with a minimal effect size (Cohen's D = 0.041). This suggests that ability to accommodate does not reliably affect the width of the zone of good stereo. The statistical power ($1 - \beta$ error probability) was approximately 0.053. The calculated power of approximately 5.28% suggests that our study had limited statistical power to detect an effect size (Cohen's D = 0.041) similar to the observed non-centrality parameter. This further accentuates the possibility that the non-significant result in the t-test may have been influenced by the study's limited power to detect potential effects.

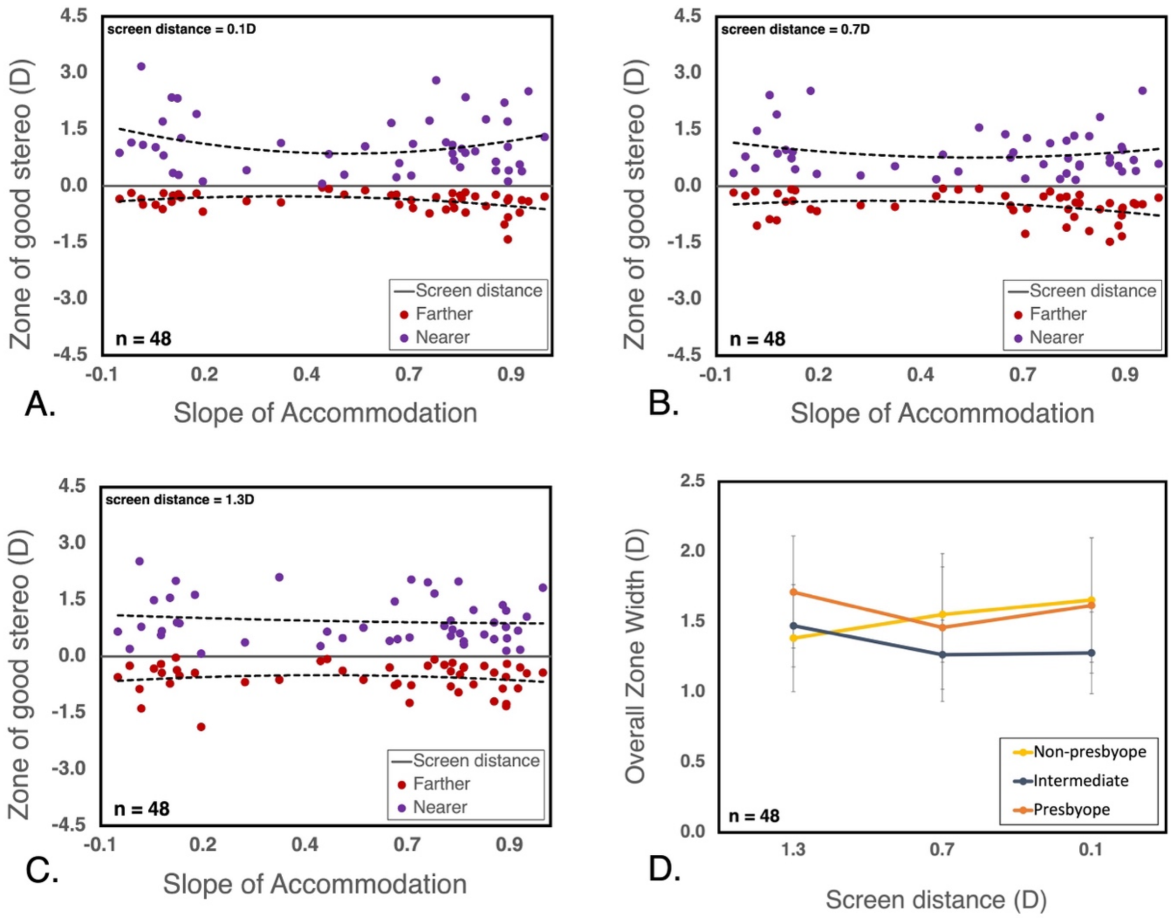



**Figure 2.11**: Relationship between zone of good stereo depth perception and slope of accommodation response. Figures in panel A, B, and C, plots individual observer's zone of good stereo as a function of their slope of accommodation response at three screen distances 1.3D, 0.7D, and 0.1D respectively. Individual datapoints in violet represent zone nearer to the screen (Positive zone) and data points in maroon represents zone farther from the screen (Negative zone). The dark black line indicates screen distance. The curves (dashed lines) are second-order polynomial fits to the data for the near and far zones. **D.** The overall zone width averaged across all observers is plotted as a function of screen distance for three presbyopia groups. The non-presbyopes, intermediates, and presbyopes are groups are indicated by yellow, dark grey, and orange lines. The error bars are ± 1 standard error of the mean.

Similar to above, we also analysed the position of the zone of good stereo relative to screen distance as a function of ability to accommodate (Figure 2.12). The pattern of results was similar to that found when the data were grouped by age, but differed in detail. The zone centres were again always nearer than the screen, and tended to become increasingly near (indicating increasing asymmetry of the zone). Moreover, while zones were overall nearest for the presbyope group, the pattern across groups was not so systematic as for the age-group analysis. A two-way (screen distance x accommodation slope) mixed-measures ANOVA found a significant main effect of screen distance ($F(2,88) = 8.57$, $p < 0.001$), but the main effect of accommodation slope was not significant ($F(2,44) = 1.21$, $p = 0.309$). The effect was minimal with a partial eta squared of 0.163 and 0.052 for screen distance and accommodation slope respectively. The screen distance x accommodation slope was also not significant ($F(4,88) = 0.74$, $p = 0.569$) with a minimal effect size (partial eta squared value of 0.032). Post hoc Bonferroni-corrected pairwise comparisons for the effect of distance showed that the zone centre was significantly nearer at 0.1 D than at both 1.3 and 0.7 D ($p = 0.001$ and 0.003, respectively). The zone positions for 1.3 and 0.7 D distances were not significantly different ($p = 1.0$).

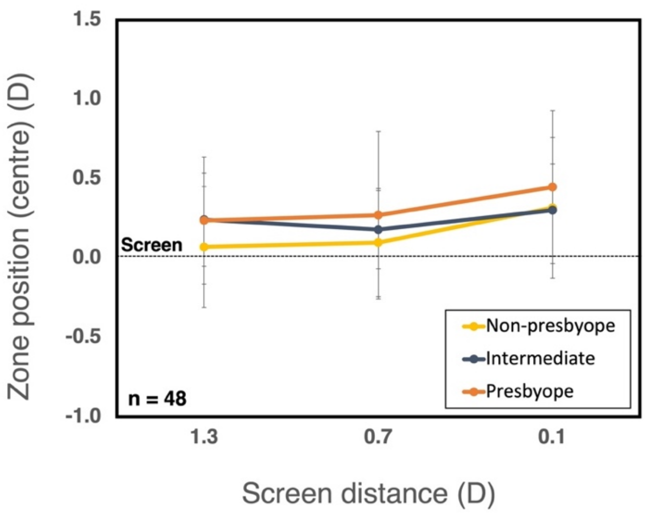



**Figure 2.12**: Relationship between position of zone of good stereo depth perception and screen distance for presbyopia groups. Zone centre positions averages across all observers is plotted as a function of screen distance from three presbyopia groups. The thin dashed line represents the screen. The non-presbyopes, intermediates, and presbyopes are groups are indicated by yellow, dark grey, and orange lines. The error bars are ±1 standard error of the mean.

*Predicting individual differences in the zone of good stereo from other visual-system characteristics*

The above analysis suggests that the overall width of the zone of good stereo is not reliably influenced by age, or the ability to accommodate. Said another way, our data do not resemble patterns expected under either of the two contrasting predictions outlined in the Introduction. We did find an effect of age on the position of the zone, however, as well as considerable individual variability in width and position of the zone of good stereo. In this section we explore whether these individual differences in the tolerance of stereo performance to vergence-accommodation conflict are related to observers' phoria, and zone of clear, single binocular vision (ZCSBV).

As discussed in the Introduction, phoria might be expected to predict the position of the zone of good stereo, because in-principle it represents the minimum-effort-state of vergence for a given screen distance. Figure 2.13 (D), plots average phoria broken down by age group. Overall, the data follow the typical pattern of esophoria at far distance (converging nearer than the accommodation distance), crossing over to exophoria at nearer distance (converging farther than the accommodation distance) (Tait, 1951; Freier and Pickwell, 1983; Shibata et al., 2011; Leat et al., 2013; Abraham et al., 2015). It can be seen that there was also a clear effect of age, with observers consistently converging nearer with increasing age at all three accommodation (screen) distances. To analyse these effects statistically we first transformed the phoria data into measurements relative to the screen distance (so they could be compared in terms of degree of esophoria/exophoria at the different distances), and then carried out a two-way (screen distance x age) ANOVA on the resulting data. There was a significant main effect of both screen distance ($F(2,94) = 103.39$, $p < 0.001$) with a large effect size (partial eta squared value of 0.687) and age group ($F(2,47) = 3.53$, $p = 0.037$) with a minimal effect size (partial eta squared value of 0.131). The screen distance x age group interaction was not significant ($F(4,94) = 0.54$, $p = 0.710$) with a minimal effect size (partial eta squared value of 0.022). For the effect of distance, all post hoc Bonferroni-corrected pairwise comparisons (1.3 vs 0.7 D, 0.7 vs 0.1 D and 1.3 vs 0.1 D) were statistically significant ($p < 0.001$), indicating that phoria was systematically nearer, with respect to the screen, at farther screen distances. Post hoc tests of the effect of age showed that phoria was significantly nearer for the older age group compared to younger (i.e. the older group was more esophoric; p = 0.036). Phorias for the young and middle age groups were not significantly different ($p = 1.0$), nor was the difference between middle and older age groups ($p = 0.191$). While these data indicate that age affects phoria, they are inconsistent with previous findings that younger individuals (age <15 years) tend to be more esophoric (vergence distance is less than accommodation distance) and older adults more exophoric (vergence distance greater than accommodation distance) by the age of about 60 years (Freier and Pickwell, 1983; Leat et al., 2013; Abraham et al., 2015).

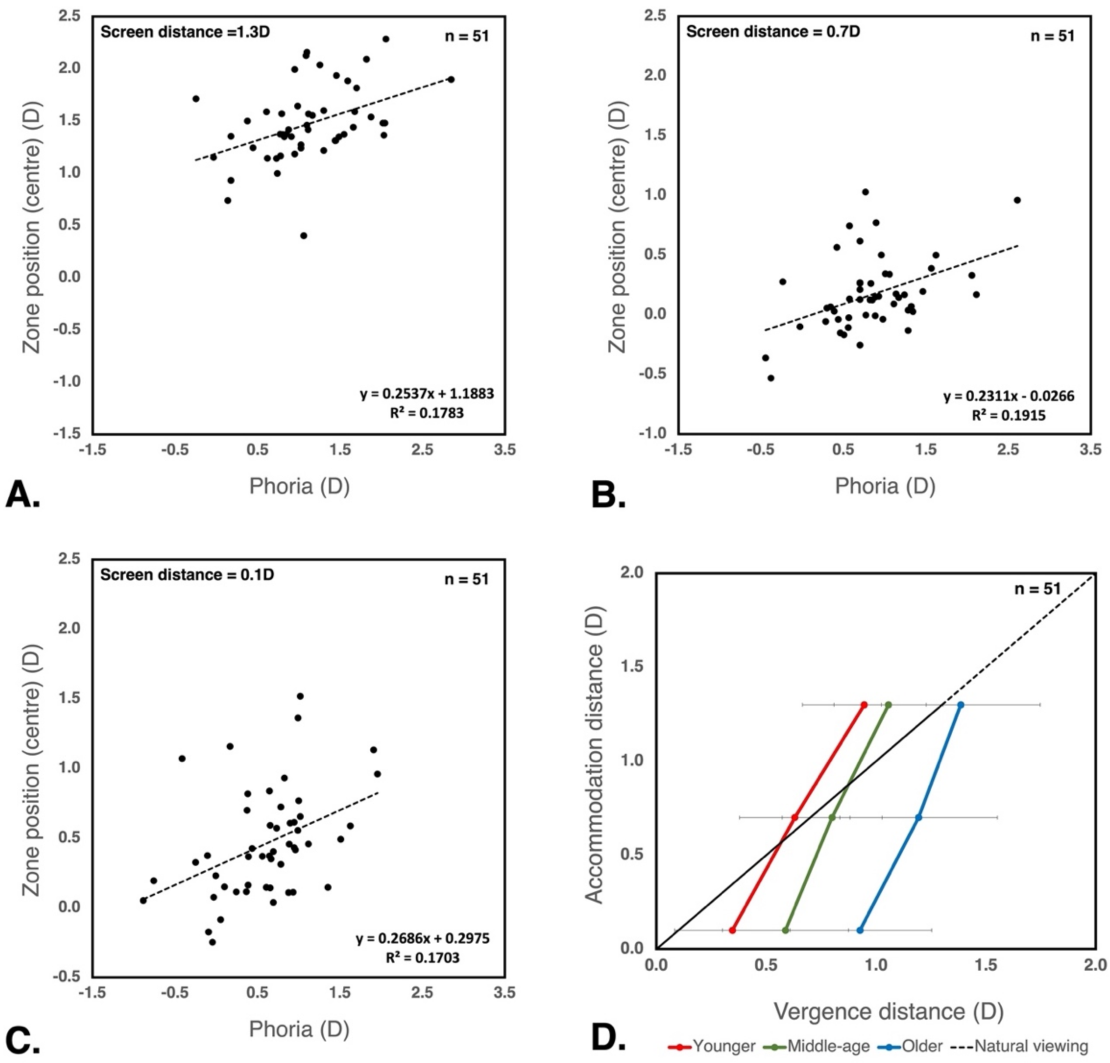



**Figure 2.13**: Relationship between zone of good stereo depth perception and phoria. Figures in panel A, B, and C, plots individual observer's centre position of zone of good stereo as a function of their phoria for three screen distances 1.3D, 0.7D, and 0.1D respectively. Pearson's $R^2$ for A (0.18), B (0.19), and C (0.17). **D**. Average phoria plotted for each age group. The vergence distance in dioptres on the x-axis and the focal distance in dioptres on the y-axis. The younger, middle-age, and older are groups are indicated by red, green, and blur lines. Individuals with 0 phoria will have a slope of 1 (shown by solid black line). The error bars are ±1 standard error of the mean.

The pattern of effects of age group and screen/accommodation distance on phoria is similar to that observed for the position of the zone centres (Figure 2.10). This is consistent with the idea that phoria predicts the position of the zone of good stereo with respect to the screen. As discussed earlier, a similar relationship was found by Shibata et al. (2011) between phoria and the zone of comfort. To investigate the ability of phoria for the zone of good stereo at the level of individual observers we next examined the correlation between individual phorias and zone positions (zone centres). Figure 2.13 (A-C) plots zone centre positions for all observers as a function of their phoria (each panels represents one screen distance). The data are plotted in units of absolute distance (in dioptres). There was a statistically significant positive correlation between phoria and the position of the zone of good stereo at 1.3, 0.7 and 0.1 D screen distance ($p = 0.002$, $p = 0.001$ and $p=0.003$, respectively). The relatively low R-squared values of the regression fits (see individual plots) show that the relationship is quite noisy. Nonetheless, the position of individual's zone of good stereo is partly predicted by their phoria, similar to the findings for the zone of comfort reported by Shibata et al. (2011).

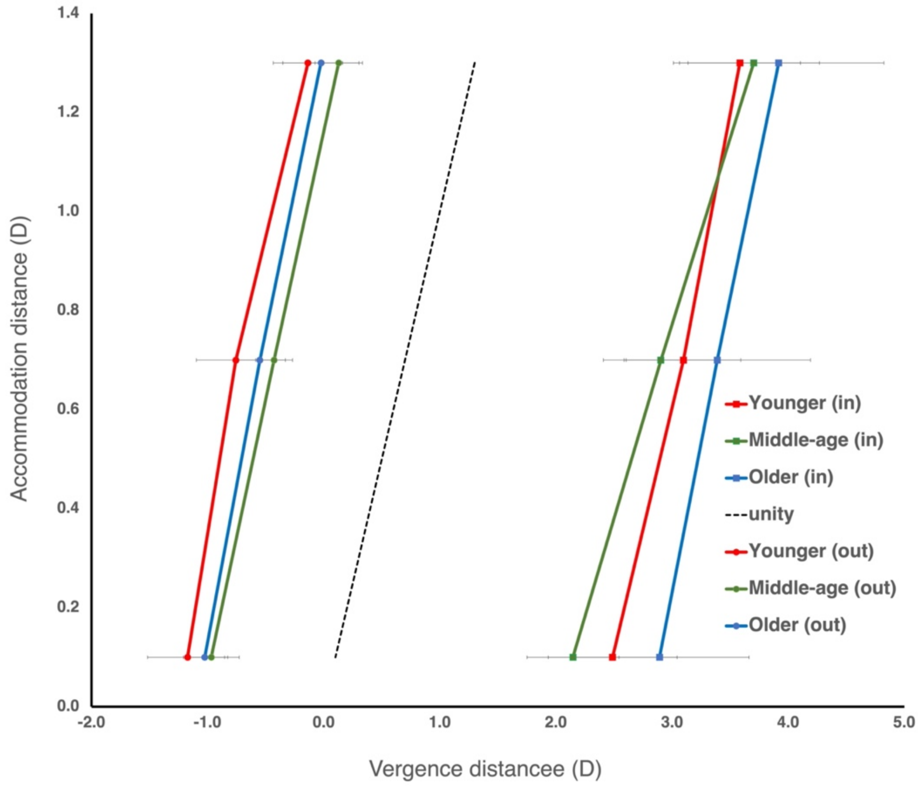



**Figure 2.14**: The estimated zone of single clear binocular vision (ZCSBV). The x-axis is the vergence distance in dioptres. The colours of the lines represent the younger (red), middle-age (green), and older (blue), which forms the average negative (outer limit) and positive (inner limit) of the ZCSBV (solid lines). This thin black dashed line is the natural viewing line accommodation and vergence responses match. The error bars are ±1 standard error of the mean.

Finally, we also examined the ZCSBV for each age group, and explored whether the width of the ZCSBV (which reflects a sort of upper limit on the ability to decouple accommodation and vergence) predicts the width of the zone of good stereo. Figure 2.14 plots the inward and outward boundaries of the ZCSBV for each age group. The outward values are primarily constrained by the ability to diverge the eyes and, as expected, they were similar across age groups. For the inward direction there is some evidence that the ZCSBV boundary is nearer for the older age group, but there is no systematic ordering with increasing age of the group. To analyse these data statistically we analysed the size of only the inward ZCSBV, relative to the screen distance, using a two-way (screen distance x age group) ANOVA. Neither the main effect of distance ($F(2,96) = 0.005$, $p = 0.995$) nor of age group ($F(2,48) = 1.05$, $p = 0.358$) was significant. The effect sizes were minimal with a partial eta squared value of 0.000 and 0.042 for effect of screen distance and age group respectively. We also explored whether there is a relationship between the width of an individual's ZCSBV and the width of their zone of good stereo. We first subtracted each observer's outward values from their inward values at each distance to give the width of the ZCSBV in each case, before computing the correlation between width of the ZCSBV and width of the zone of good stereo. The data are plotted in Figure 2.15, panels A – C. It can be seen that there is no evidence of a systematic relationship between ZCSBV and width of the zone of good stereo. All three correlations were non-significant ($p > 0.5$). Taken together, we found no reliable relationship between the ZCSBV and the zone of good stereo performance.

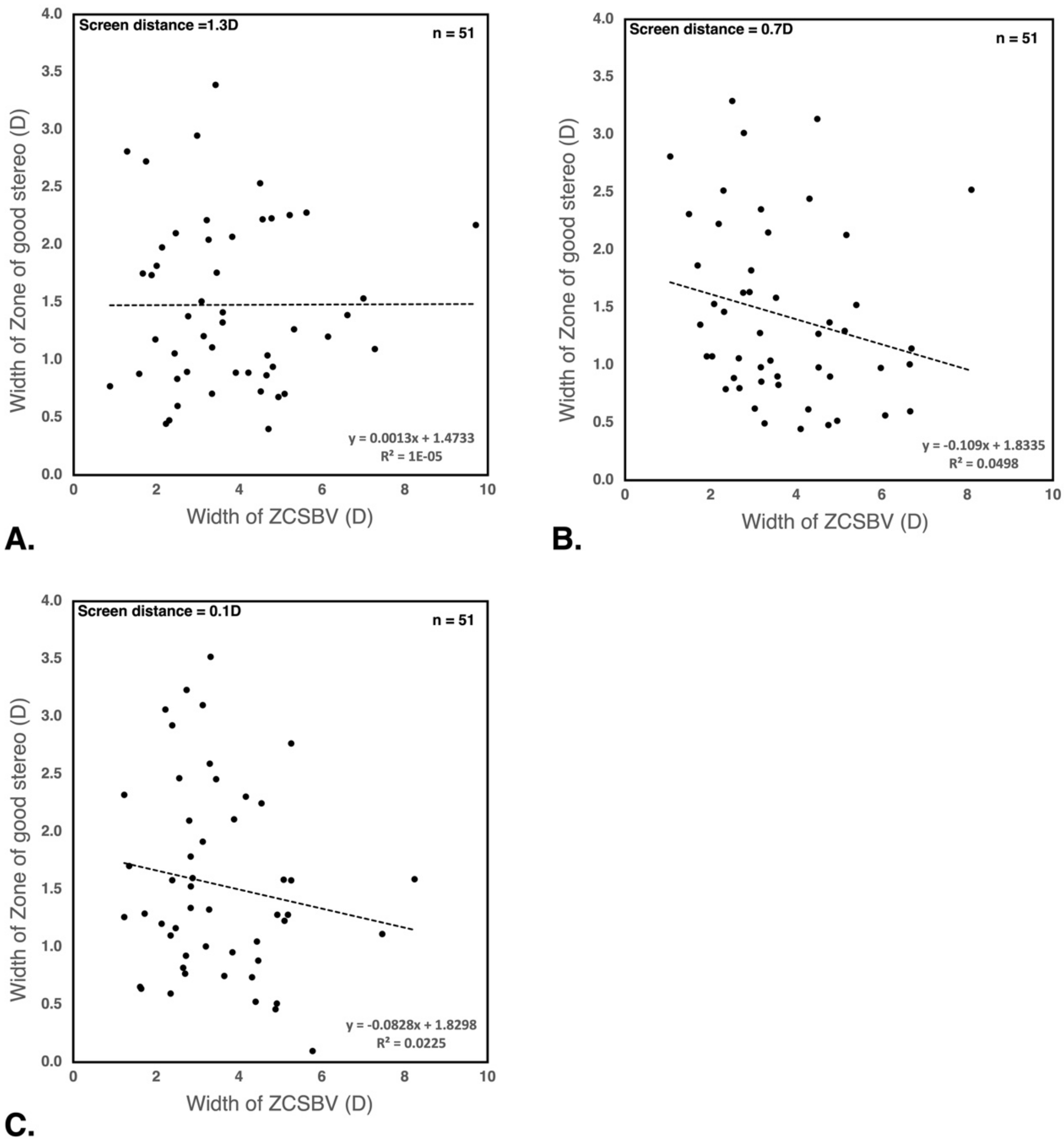



**Figure 2.15**: Relationship between zone of good stereo depth perception and zone of single clear binocular vision (ZCSBV). Figures in panel A, B, and C, plots individual observer's width of zone of good stereo as a function of the width of their ZCSBV for three screen distances 1.3D, 0.7D, and 0.1D respectively. Pearson's $R^2$ for A ($1x\ 10^{-5}$), B (0.05), and C (0.02).

## 2.4 Discussion

In our study we aimed to establish individuals' tolerance of stereoscopic depth perception to vergence-accommodation conflict, and find factors that would predict this tolerance. These data could then potentially be used manage and mitigate effects of vergence-accommodation conflict in conventional displays. We did this by establishing a range of stereoscopic depths around the screen plane (range of vergence-accommodation conflicts) that results in a given level of degradation of depth perception. The main predictor(s) of our concern was individuals' age, and age-related changes in their ability to accommodate. We considered two theories. First, the idea that younger viewers will be more robust to the effort required to decouple vergence-accommodation conflicts, so will be more tolerant than older people to vergence-accommodation conflict (Mendiburu, 2009; Banks et al., 2012). Second, older people, with reduced ability to accommodate (i.e. presbyopes) might be more tolerant to vergence-accommodation conflict because the requirement for viewing stereoscopic 3-D matches their normal pattern in real-world viewing of varying vergence with fixed accommodation (Yang et al., 2011; Banks et al., 2012).

*Age and ability to accommodate do not predict the size of individuals' zones of good stereo depth perception*

We found no reliable relationship between the width of the zone of good stereo and either age, or degree of presbyopia. It is possible that the effects predicted by both theories are present, but largely cancel each other out. That is, people become less tolerant to exerting effort (to decouple accommodation and vergence) with age, but simultaneously are required to exert less effort as they become more presbyopic. This would not be evident in mean zone-width measurements, but may manifest as increased variability at each 'end' of our sample, due to individual differences in how much each factor is in play. It is unlikely that our sample size is sufficient to detect changes in variance with age. Visual inspection of Figure 2.9. does not suggest any such effect. However, there is some evidence for increased variance across particip-ants either end of the spectrum of ability to accommodate (Figure 2.11). As discussed in the Introduction, age is not a perfect predictor of ability to accommo-date, and so it is important to analyse the zone-width data according to both age and ability to accommodate in order to evaluate the theories. Consistent with this, our results showed that there is a high degree of variability in ability to accommodate in observers between the age ~30–50 years, presumably due to variability in onset of presbyopia. Yet, the fact that neither variable predicted the width of the zone of good stereo suggests that as people age, and become more presbyopic, their overall tolerance of stereo depth perception to accommodation-vergence conflicts does not change. This suggests that spatial stereo 3-D content cannot usefully be tailored to viewers of different ages with a view to optimising stereo depth perception.

An exception to this is that we did observe an effect of age (but not of ability to accommodate) on the position, or asymmetry with respect to the screen, of the zone of good stereo, with older observers showing nearer zones than younger observers at all three focal distances. We did not predict this pattern of effects, and it is unclear what the causal mechanism might be. One possibility is that older adults have both a limited rage of accommodation, and that their resting point of accommodation is nearer, so that effectively they experience smaller accommodation-vergence conflicts nearer than the screen. We also found a relationship between zone position and phoria, which we discuss below.

*Other predictors (Phoria and ZCSBV)*

Based on theoretical considerations, and previous research (Shibata et al., 2011), we expected that phoria and the zone of single clear binocular vision (ZCSBV) might also predict the zone of good stereo depth perception. We found some evidence that phoria is related to the zone of good stereo. Specifically, we found that the phoria was nearer with increasing age, as observed for the position of the zone of good stereo. Moreover, when we found that the position of individuals' zones of good stereo correlated significantly with their phorias. This finding is consistent with Shibata et al.'s (2011) finding that phoria was predictive of the zone of comfort position. However, the pattern of changes in phorias that we observed

— increasingly nearer with increasing age—is not typical of that reported previously in the literature. The general consensus from previous studies is that average phoria for near distance increases with age from early 20's, in a steady progression, becoming exophoria by the age of 60 years (Freier and Pickwell, 1983; Leat et al., 2013; Abraham et al., 2015). In our data older individuals were more esophoric (vergence nearer than accommodation distance) than the younger group. It should be noted that most of these studies evaluated phoria at accommodation distances nearer than ours (a Maddox Wing is used at 33 cm or 3 D), whereas our 'near' distance was 1.3 D, and in general (and in our data) people are more esophoric at farther distances. It seems unlikely, however, that this can entirely account for the direction of the difference we observed across ages, which remains puzzling.

As noted above, our results showed a statistically significant positive correlation between phoria and the position of the zone of good stereo at all three screen distances. Shibata et al., (2011) found a similar relationship between phoria and zone of comfort. The authors suggested that the correlations were low because they used questionnaire data for discomfort, which is inherently noisy. Our measure of the zone of good stereo is, in principle, relatively precise, yet we too observed quite low correlations with phoria. Taken together, this suggests that while phoria is clearly a determinant of tolerance to vergence-accommodation conflict, other significant factors must also be involved.

We did not find any reliable relationship between the ZCSBV and the zone of good stereo depth perception. The ZCSBV is a measure of the upper limit of an individual's ability to decouple the accommodation and vergence responses, and so might reasonably be expected to relate to the zone of good stereo. Consistent with this Shibata et al. (2011) found that the ZCSBV did predict the size of the zone of comfort. We did not find such a relationship, however. Nor did we find any effect of distance and age group on individuals' ZCSBV.

*Comparing the zone of good stereo to the zone of comfort*

The underlying factors contributing to the phenomena of the “zone of comfort” and the “zone of good stereo depth perception” are intrinsically linked, as they are primarily determined by an individual’s capacity to decouple the neural coupling between vergence and accommodation. The degree of difficulty in decoupling this neural link directly impacts the exertion of effort by the visual system and the accuracy of oculomotor responses. Consequently, it seems likely that both the “zone of comfort” and the “zone of good stereo depth perception” are related to the zone of single clear binocular vision (ZCSBV) in similar ways. The ZCSBV measures an individual’s maximum ability to decouple vergence and accommodation (i.e. when trying hard to maintain clear, single binocular vision). Under normal viewing, the effort required to decouple vergence and accommodation is likely to cause noticeable discomfort, and impair stereo depth perception, at smaller conflicts than the maximum possible decoupling (indicated by the ZCSBV). this means that the width of both zones would be expected to be narrower than the ZCSBV. Discomfort and stereo depth perception are also likely to interact, however. For example, exerting more effort to decouple vergence and accommodation might result in better stereo depth perception, but also increased discomfort. In experiments, we do not control (nor is it obvious how we could) the level of effort people exert, and so it is difficult to understand whether the zone of comfort is meaningfully wider or narrower than the zone of comfort. To investigate the potential overlap between the zone of comfort and the zone of good stereo depth perception, we plotted the boundaries (the vergence distances) of the measures of these zones as a function of focal distances (Figure 5.1). The estimates for the far and near boundaries of the zone of comfort were derived using equation 7 from Shibata et al. (2011). The width of the zone of good stereo depth perception exceeded that of Shibata's zone of comfort. However, it is important to note that in both cases the ‘boundaries’ of the zones are arbitrary, in that they reflect a criterion level of discomfort rating, or stereoacuity decrement. Moreover, they presumably reflect a continuous underlying distribution rather than discrete zones. Direct comparison between the quantitative data for the zone of comfort and the zone of good stereo depth perception is not feasible, and any conclusions drawn remain speculative.

*Implications*

One purpose of our study was to provide guidelines on optimising stereo 3-D content across a wide range of audiences, if possible, by using 'observable' demographic information such as age. The motivation for our study comes from the fact that presenting focus cues correctly is difficult, and involves development and use of complex hardware and software, and the cost of replacement of existing display technologies could be prohibitive. This means that for some applications conventional stereo 3-D may remain an effective option, in which case understanding how to minimise the adverse effects of incorrect focus cues remains important. Factors such as age, ability to accommodate, phoria, and ZCSBV have the potential to affect tolerance to conflicts, and if these properties are known, or can be predicted, for a target audience optimisation of 3-D content is possible. Given that no relationship was found between people's zone of good stereo depth perception and their age or ability to accommodate, this seems to be challenging. Phoria on the other hand seems to partially predict the zone centre positions, similar to what Shibata et al. (2011) found. This is not a very practical measure on which to tailor content, however, given that most people do not know their phoria and the relationship between phoria and the zone centre is quite noisy.

Despite this lack of effects of age/accommodation, our data nonetheless provide information about the range of vergence-accommodation conflicts over which stereo performance remains at a reasonable level. As such, the data are complementary to the findings for the zone of comfort (Shibata et al., 2011), providing insight into a different yet important aspect of stereoscopic 3-D displays that should be considered in evaluating their effectiveness. Moreover, the overall consistency of our findings (with some exceptions) to those for the zone of comfort provide corroborating evidence, from a larger, more diverse sample, of the idea that there is a measurable tolerance to vergence-accommodation conflict, reported by Shibata et al. (2011), and that this has properties that can be used to generate guidelines for content producers. That is, we can be more confident in the recommendations Shibata et al. (2011) made on the basis of their study.

*Limitations and future research*

Although we have shown that a zone of good stereo can be determined empirically, it is important to recognise that it is non-trivial to translate these findings to predictions for real-world viewing, for several reasons. First, the size of our measured zone is arbitrary, in that it depends on an arbitrarily chosen change in stereo performance (from 90% to 75% correct on our particular task). While this results in comparable results across observers and distances it means the absolute size of the zone is not necessarily all that meaningful. Note, a similar argument can be made for Shibata et al.'s (2011) zone estimates. Relatedly, the zone of good stereo (and comfort) is not discrete, but is continuous, with performance deteriorating gradually, from smaller conflicts than the measured zone boundary, and continuing beyond.

Second, the functional consequences of poorer stereoacuity for different real-world tasks, or for the appearance of scenes in the case of entertainment applications, is not well understood. It is therefore difficult to make specific predictions about precisely how these aspects of the stereo 3-D experience will be affected by vergence-accommodation conflicts.

Third, our study examined the effects of spatial aspects of vergence-accommodation conflicts for stereo performance. Practical applications of stereoscopic 3-D also often involve varying temporal structure to the vergence-accommodation conflicts (consider scene-cuts in movies, for example, or viewing static vs rapidly moving objects in stereo 3-D). It has been shown that rapidly varying vergence–accommodation conflicts result in more visual discomfort than slowly varying conflicts (Kim et al., 2014), and that the time required to fuse stereoscopic images increases with magnitude of conflict. It therefore seems likely that the temporal structure of vergence-accommodation conflicts may affect stereo performance and should form part of the consideration for producing 3-D content. Indeed, it is possible that the effect of the speed of vergence-accommodation conflict on stereo performance could reveal effects of age and ability to accommodate that were not present in our study.

Fourth, it remains possible that discomfort and fatigue *are* affected by age-related changes in the ability to accommodate, even though stereo performance is not. Although the pattern of our results is broadly analogous with the zone-of-comfort findings of Shibata et al. (2011) it does differ in some ways (the lack of predictive power of the ZCSBV, for example), suggesting that the two measures are not entirely correlated. Moreover, the idea that young people may be more tolerant to the effort required to decouple accommodation and vergence (Mendiburu, 2009) arguably relates to robustness to the effort required, which could conceivably affect the subjective sense of discomfort more than it affects stereo performance. Said another way, younger and older people could be equally able to decouple their vergence and accommodation responses, but it could be more uncomfortable for one or other group.

Overall, our data suggest that age-related changes in the ability to accommodate do not predict tolerance of stereo performance to vergence-accommodation conflict, and so do not support either of the two opposing theories in the literature about which demographics should be most (or least) affected by conventional stereo 3d presentation.

# Chapter 3

## 3 Correct focus cues improve perceived realism

### 3.1 Introduction

Stereoscopic 3d imagery often aims not only to create a given quantitative percept of 3d scene structure, but also to convey a qualitative appearance of depth that is as realistic and natural as possible. Advances in photorealistic rendering techniques, alongside improvements in display hardware, mean that most cues to 3d scene structure can now be presented with sufficient fidelity that the 'gap' to real scenes is small (de Silva et al., 2011; Banks et al., 2016; Zhong et al., 2021). Focus cues remain an exception, however. Stereo 3d displays still typically present images on a single, fixed display surface resulting in an incorrect stimulus to the eye's focusing response (accommodation) and incorrect depth-dependent retinal blur. This study investigates whether presenting correct focus cues increases the realism and naturalness of perceived depth in 3d scenes—referred to as *depth realism* (Hibbard et al. 2017)—compared to conventional 3d presentation in which focus cues are incorrect. The results contribute to fundamental understanding of the role of focus cues in scene perception. Moreover, approximating correct focus cues is technically challenging, increases the cost and complexity of display systems, and can require trading-off other desirable aspects of image quality (Lanman & Luebke, 2013; Huang et al., 2015; Chang et al., 2018; Javidi et al., 2021; Chakravartula et al., 2022). It is useful therefore to evaluate the potential benefits that correct focus cues can confer, in order to inform the development track of future display technologies.

In recent years the quality that can be achieved in stereo 3d imagery has improved significantly due to improvements in areas such as the spatial resolution and dynamic range of display technologies, combined with continued developments in photorealistic rendering techniques (Gortler et al., 1996; Pereira et al., 2021; Hu & Hua, 2014; Zhong et al., 2021). Presenting focus cues correctly remains challenging, however. Most if not all current stereo 3d display products still present images 'conventionally', using a single display surface at a fixed distance. Consequently, focus cues resulting from natural viewing, and from viewing an otherwise equivalent scene depicted in stereo 3d, differ substantially (Mather & Smith, 2000; Akeley et al., 2004; Hoffman et al., 2008; Watt et al., 2005). The stimulus to accommodation does not vary naturally with variations in scene depth but is instead consistent with the display surface. Moreover, the patterns of defocus blur that naturally result when scene points are nearer or farther than the currently focused distance are not reproduced correctly.

Research into the consequences of incorrect focus cues has mostly concentrated on the unnatural demands placed on the oculomotor system by stereo 3d viewing. The mismatch between the stimulus to binocular vergence eye movements and the (fixed) stimulus to accommodation gives rise to so-called vergence-accommodation conflicts. Accommodation and vergence responses are neurally coupled, via vergence-accommodation cross-links (Fincham & Walton, 1957; Martens & Olge, 1959; Schor, 1992), and the effort required to decouple them is a primary cause of discomfort and fatigue in stereo 3d viewing (Wann et al., 1995; Emoto et al., 2005; Okada et al., 2006; Hoffman et al., 2008; Lambooij et al., 2009; Shibata et al., 2011). Moreover, because decoupling may not always be possible, vergence and/or accommodation may be inaccurate (Emoto et al, 2005; Watt et al., 2005; Willemsen et al., 2007), causing degraded stereo depth perception (reduced stereoacuity and slower stereoscopic fusion; Watt et al., 2005; Hoffman et al., 2008). The consequences of incorrect focus for the subjective appearance of 3d imagery, and specifically for depth realism, remain unclear, however.

Incorrect focus cues might be expected to reduce depth realism for several reasons. One reason is that they provide inaccurate depth information, which could result in distortions in perceived depth, and discernible conflicts between depth cues, that do not occur in natural viewing. Although focus cues have often been thought of as 'weak' depth cues, they can in fact play a significant role in depth perception under some circumstances. Varying accommodation distance has been shown to affect perceived slant of stereoscopically defined surfaces, for example, by altering the estimate of distance used to interpret binocular disparities (Watt et al., 2005; Hoffman et al., 2008). Incorrect retinal blur gradient has been found to have little or no effect on the perceived slant of fixated, isolated surfaces (Watt et al., 2005a, 2005b). However, blur contributes significantly to depth perception at occlusion edges (Marshal et al., 1996; Mather, 1997), and for scene points nearer and farther than fixation (Kumar & Glaser, 1992; Held et al., 2012). Moreover, the phenomenon of tilt-shift miniaturisation, in which increasing the 'global' blur gradient causes natural scenes to resemble scale models, demonstrates that blur can dramatically affect perception of overall spatial scale (Laforet, 2007; Held et al., 2010; Vishwanath & Blaser, 2010). Incorrect focus cues may therefore reduce realism by providing quantitatively incorrect information about depth.

Incorrect focus cues could also diminish the qualitative sense of three-dimensionality present when viewing the real world. Viewing real 3d scenes results in a qualitatively distinct sense of vivid, tangible, 'real' 3d structure, referred to as *stereopsis,* which is typically not present when viewing 2d paintings, or even photographs (Ponce & Born, 2008; Vishwanath et al., 2013). It has long been recognised that this sense of stereopsis is induced by viewing 3d scenes defined solely by binocular disparity or motion parallax (Rogers & Graham, 1979, 1982; Braunstein, 1966, 1968, 1976). More recent work suggests that relative blur can also induce the sense of 'real depth' associated with stereopsis (Vishwanath & Hibbard, 2013). Incorrect blur in stereo 3d might therefore be expected to diminish the sense of stereopsis, reducing depth realism, compared to when focus cues are correct.

Incorrect focus cues in stereo 3d could also diminish realism not in terms of perceived depth per se, but because the patterns of retinal blur, and associated accommodation responses, do not match our experience of viewing real-world scenes. For example, when fixating a near object, a farther object would be expected to be blurred. If we then choose to look at it, we make it sharp by accommodating, and the originally fixated object becomes blurred. These static and dynamic aspects of retinal blur, which are entirely predictable in real-world viewing, are not reproduced correctly in conventional stereo 3d (Creem-Regehr et al., 2005; Sahm et al., 2005; Willemsen et al., 2007; Backus et al., 1999; Buckley & Frisby, 1993; Ellis et al., 1993; Frisby et al., 1996; Frisby et al., 1995; Watt et al., 2005). This could provide a straightforward signal that the scene is unnatural, potentially reducing depth realism. Moreover, motor output itself could directly cue the unnatural nature of the scene. That is, viewers may detect that they are not making the normal, expected pattern of accommodation responses as they look around the scene, and/or that greater than normal motoric effort is required, again reducing resulting realism and naturalness of stereo 3d imagery.

Although previous work on depth realism in stereo 3d has not directly examined the impact of focus cues, it does provide some hints that incorrect focus may reduce perceived realism. In a study by Hibbard et al. (2017), observers viewed two pairs of random-dot-defined planes, separated in stereoscopic depth by varying amounts, and presented using conventional stereo 3d (i.e. with incorrect focus cues). In separate sessions observers made two-alternative, forced-choice judgements either of which pair had the *largest* depth separation, or which had the most *realistic* depth separation. All possible combinations of depth separations were presented and Thurstonian scaling was used to provide relative measures of perceived depth magnitude, and depth realism. As expected, the magnitude of perceived depth separation increased with increasing disparity-specified depth (levelling off around the limit of binocular fusion). Depth realism showed a different pattern, however. Realism was highest at small depth separations and reduced systematically as separation-in-depth increased. While these results demonstrate that magnitude and realism are dissociable aspects of perceived depth, they are also puzzling. One possible explanation, described by Hibbard et al. (2017), is that depth realism depends on the precision of the available depth information, with more precise depth estimates appearing more realistic. Depth separations are less precisely encoded as separation-in-depth increases, as evidenced by increasing discrimination thresholds (see Hibbard et al.'s Experiment 2). This change in precision is not unique to stereoscopic imagery, however. Depth information also becomes less precise with increasing separation-in-depth in real scenes, largely for geometrical reasons (Hillis et al., 2004; Keefe et al., 2011). Yet, it seems implausible that the real world would appear less realistic as separation between objects increases. An alternative explanation is that Hibbard et al.'s (2017) data reflect the influence of incorrect focus cues. The stimuli were presented on a single display surface, and so the magnitude of the error in focus-cue presentation (i.e. the mismatch between depth separation specified by focus cues and by disparity) became larger as depth separation increased. It is possible that the decreasing judgements of depth realism reflected this increasing focus-cue error. Consistent with this, the roll-off in depth realism observed by Hibbard et al. (2017) occurred at a disparity-specified separation corresponding to between ~0.15 and 0.3 dioptres (D), which is similar to the functional depth-of-focus of the eye (Campbell, 1959; Green et al., 1980). Thus, depth realism began to decrease at the point at which relative blur in the non-fixated stimulus plane would just be detectable in real-world viewing.

In the current study we examine whether incorrect focus cues reduce the realism and naturalness of perceived depth. Our method was similar to Hibbard et al.'s (2017) study but included direct comparisons of stimuli presented with correct versus incorrect focus cues (while holding all other aspects constant). We used disparity-defined stimuli (random-dot stereograms) rather than images of real scenes in order to remove uncontrolled variables that could otherwise affect depth realism. Correct focus cues were presented using a multiple-focal-planes stereoscopic display

(MacKenzie et al., 2012; Watt et al., 2012), which creates a stack of transparent, fronto-parallel image planes at different focal distances. Using this display technique, continuous variations in focal distance can normally only be approximated, by interpolating between image planes (Akeley et al., 2004; MacKenzie et al., 2010). Here, we used planar stimuli (as Hibbard et al., 2017), and positioned the (movable) focal planes so that they precisely coincided with the stimulus planes, allowing us to present focus cues fully correctly. In the conventional stereo 3d conditions we presented equivalent stimuli on a single focal plane. As such, our study is not a test of the multiple-focal-planes display approach per se, but instead uses it as a tool to compare conventional stereo 3d presentation with truly correct focus cues.

## 3.2 Methods

**Observers**

Fifteen participants aged 20 - 32 years (Females: 9 and Males: 6) took part in the Experiment. All participants had normal or corrected to normal vision, and normal stereo-acuity (assessed by the Randot stereo test; Stereo Optical Company, Inc.). All participants were naïve to the purpose of the experiment. The experiment was approved by the departmental ethics committee of Bangor University and conducted in accordance with the Declaration of Helsinki.

**Apparatus and stimuli**

*Multiple-focal-planes display*

We use a multiple-focal-planes display (MacKenzie et al., 2010) (details of the display is available in section 1.7).

*General stimulus properties*

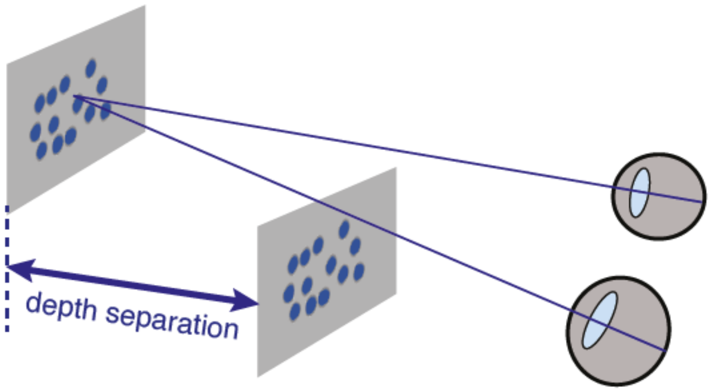



**Figure 3.1:** Test stimuli: Each stimulus interval contained two non-overlapping random-dot defined planes, separated in depth.

Each stimulus interval consisted of a pair of random-dot-defined fronto-parallel rectangles, vertically separated, and with varying depth separations between them specified by binocular disparity. The lateral separation of left- and right-eye's displays was adjusted to match each observer's interocular distance (IOD), and disparities were calculated taking each observer's IOD into account. Figure 3.1 depicts one pair rectangles in cartoon form. Stimulus size was limited by the field of view of the display, which is constrained by the physical size of the Badal lenses. The individual rectangular planes were on average 1.6 degrees high and 6.0 degrees wide. A random jitter was added in the range +/- 0.25 degrees for the height dimension, and +/- 1.5 degrees for width (drawn from a uniform distribution) to prevent relative size of the rectangles providing a reliable cue to depth separation. The vertical separation of the two rectangles was on average 0.3 degrees. The random dots were white [circular in shape, size = 2 degrees, presented with a jitter of (x = 1.5 degrees and y = 0.25 degrees)], with a dot density of 3.0 dots per degree $^{2}$, and rendered with anti-aliasing. Whether the upper or lower rectangle was nearer was chosen at random on each stimulus interval. We used vertically separated stimuli rather than overlapping 'transparent' stimuli (as used by Hibbard et al., 2017) to avoid having occlusions, which multiple-focal-planes displays do not present correctly (Narain et al., 2015).

The luminance of the dots was 15 cd/m$^2$. Luminance and white-point calibration was carried out separately for each focal plane position, in each eye's display, to take into account the effects on luminance and chromaticity both of the beam-splitter optics, and of spatial non-uniformities in the monitors. Note also that because the display uses Badal optics, the angular spatial resolution of the images is the same for all focal-plane distances. This meant that the spatial resolution of the images was identical across focal-plane separations and, critically, across stimuli presented on a single display surface (i.e conventional stereo 3-D presentation) and with correct focus cues.

*Focus cue conditions*

There were three focus-cue conditions: (i) *conventional stereo 3-D*, (ii) *correct focus cues*, and (iii) simple Gaussian *rendered blur*. The conditions are shown in schematic form in Figure 3.2. In all three conditions the near rectangle was always at the same disparity-specified and focal distance, coincident with the near focal plane on our display, positioned at 1.3 dioptres (D) (76.92 cm). Variations in depth separation between rectangles were achieved by moving the far rectangle. In the conventional-stereo-3-D condition the stimuli were presented on a single display surface—the near focal plane at 1.3 dioptres (D) (76.92 cm)—replicating conventional stereoscopic 3-D presentation. In the correct-focus-cues condition, both rectangles were presented at the correct focal distance (i.e., at a focal distance that coincided with the disparity-specified distance). This was achieved by moving the mid and far focal planes as required between blocks of trials. Figure 3.3 shows the plane positions used for an example combination of depth separations. We were concerned that a false-positive result for correct focus cues (vs. conventional stereo) could arise from observers simply responding preferentially to the presence of discernible blur in the stimulus, irrespective of depth realism. This was a particular concern because random-dot stereograms are themselves unrealistic, which could make it more difficult for observers to maintain an appropriate internal standard for judging depth realism. The rendered Gaussian blur condition was a ’foil’, designed to detect such a pattern of responses by presenting focus cues in an exaggeratedly unrealistic way. For each depth separation we applied Gaussian blur to the far rectangle creating an approximate empirical match to the real blur caused with correct focus cues when fixating the near plane. This resulted in obviously incorrect blur when observers moved fixation between near and far stimulus planes (see Procedure) because, regardless of fixation, the near plane remained sharp, and the far plane remained blurred. We reasoned that if observers showed a clear preference for obviously unrealistic blur, we could not trust that their judgements of correct-focus-cues stimuli reflected depth realism per se, and so they should be removed from the analysis (see Results). Although the rendered Gaussian blur condition does not speak to our hypotheses, we also included it in the magnitude-judgement variant of the experiment so that any difference in patterns of effects of correct focus vs. conventional stereo across magnitude and depth-realism judgements could not be due to the inclusion of different conditions. For each depth separation, the width of the Gaussian blur kernel was adjusted to approximately match, empirically, the appropriate real blur (by viewing the two stimulus types back-to-back). Thus, fixating the near rectangle resulted in approximately correct relative blur, though higher-order aspects such as effects of individual eye optics, microfluctuations in accommodation, and chromatic aberration were incorrect (Nguyen, et al., 2005; Cholewiak et al. 2017).  Because the stimuli were presented on a single focal plane, however, the stimulus to accommodation from the far rectangle was incorrect, and when fixated, the near rectangle remained sharp, and the far rectangle remained blurred.

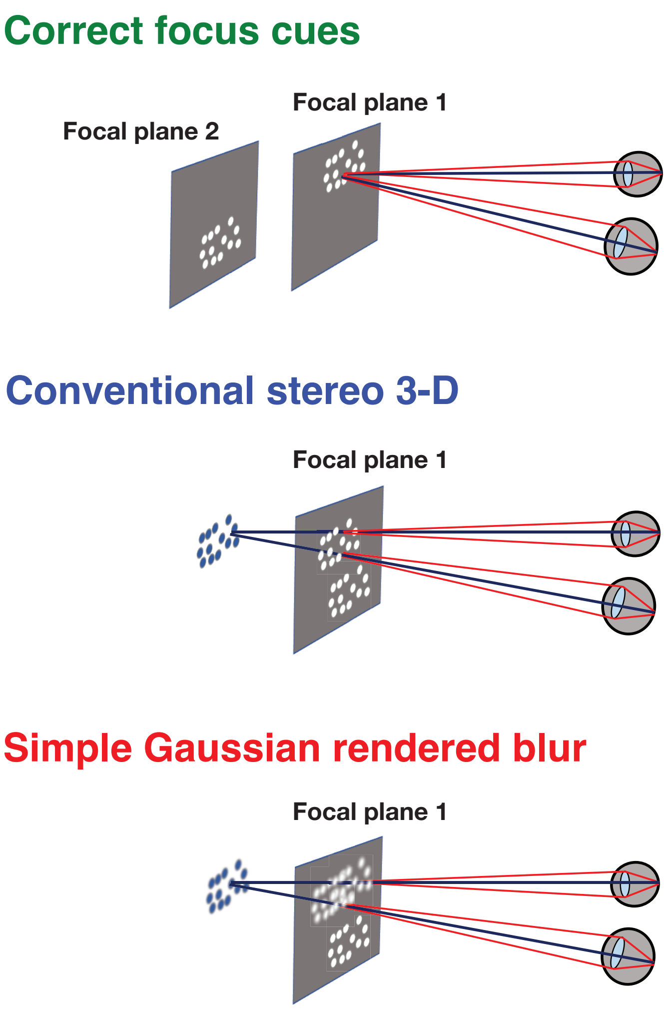



**Figure 3.2**: Focus-cue conditions: **A**. Correct focus cues condition both rectangles were presented at the correct focal distance (i.e., at a focal distance that coincided with the disparity-specified distance), this also provided correct stimulus to accommodation (red line) and vergence (blue line) distances. **B**. 3-D display (incorrect focus cues) condition both rectangles were presented on a single display surface—the near focal plane at 1.3 dioptres (D) (76.92 cm), and **C**. Simple Gaussian rendered blur condition is similar to 3-D display condition, but a simple Gaussian blur was added to the dots on the 'far' rectangle. Simple Gaussian rendered blur condition also acted as a 'foil' condition to detect if observers preferentially chose blurry stimuli, independent of correctness of focus cues.

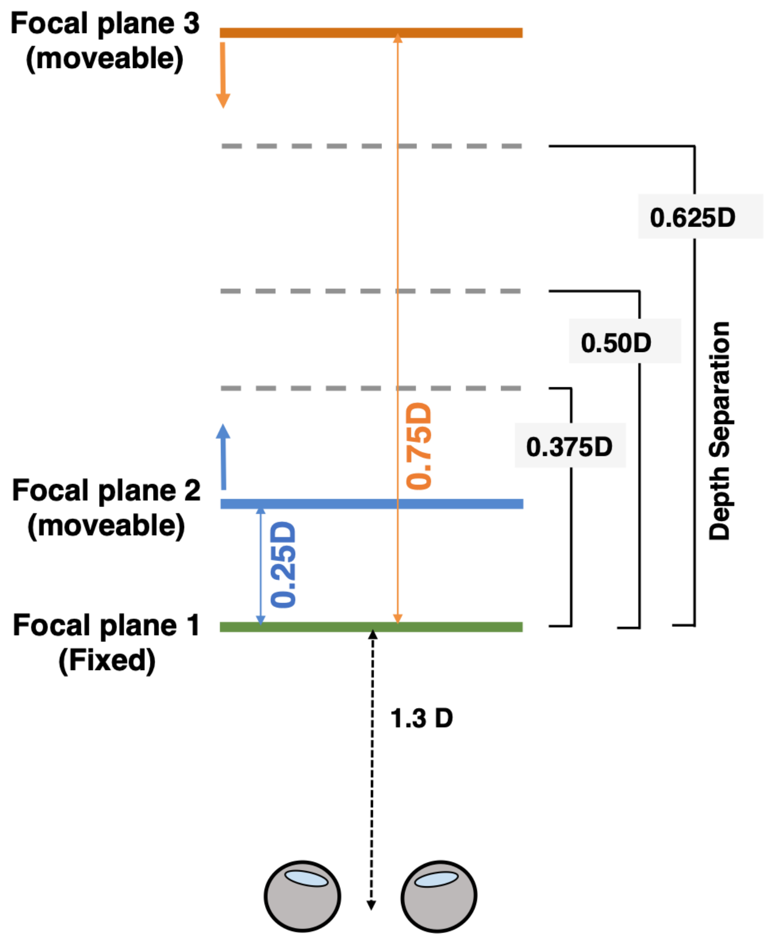



**Figure 3.3**: Plane positions for all depth separations: The figure shows an example for a comparison between 0.25D and 0.75D depth separations. The near plane (green) remained fixed at 1.3D from observer. The far planes (blue and orange) were adjusted for appropriate depth separation for depicting correct focus cue depth separation in each block.

## Procedure

Observers completed two variants of the experiment: one in which they made depth magnitude judgements, and one in which they judged depth realism. All other details of the experiment variants were identical. Observers fully completed one variant before undertaking the other. Eight observers completed the depth-magnitude variant first and seven completed the depth realism variant first.

At the beginning of each experiment block observers checked and if necessary adjusted their head position vertically and horizontally to ensure their eyes were correctly centred on each eye's display. To do this, dots were presented to both eyes on the geometric centres of each focal plane and observers adjusted their position until the three dots in each eye were superimposed. The experimenter checked that the observer's eyes were the correct distance from the Badal optics by aligning the outer (lateral) canthus—the point where upper and lower eyelids meet, which approximately corresponds to the eye's nodal point (Elliott, 2007) —with a reference marker on the apparatus.

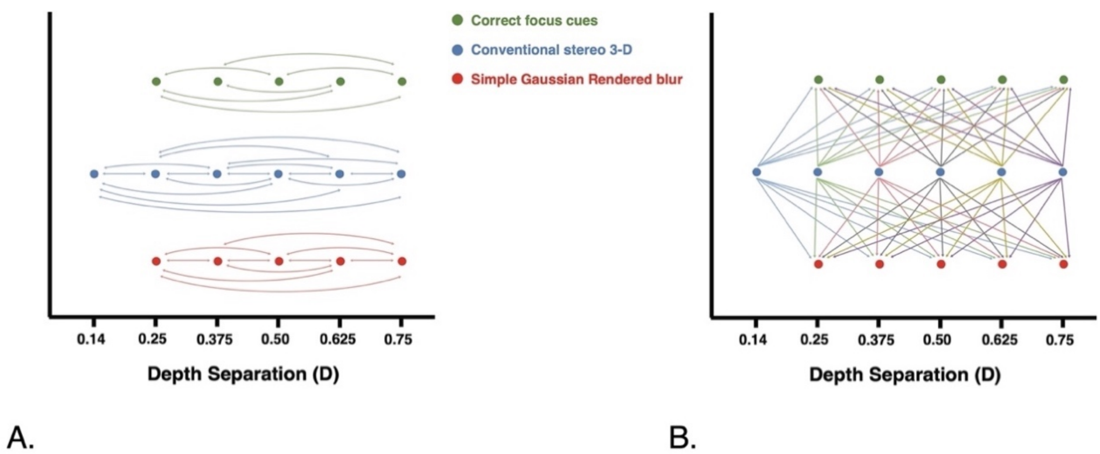



**Figure 3.4**: Stimulus pairs presented across within and across condition. **A**. Within focus-cue conditions: comparisons were made across depth separations. **B**. Across focus-cue conditions: Comparisons were made across all the three focus-cue conditions across all depth separations.

Each trial consisted of two stimulus intervals, each containing a pair of random-dot rectangles separated in depth (see *Apparatus and Stimuli*). The pairs of stimuli could be drawn from the same or different focus-cue conditions (see below). On each trial participants made a two-interval, forced-choice response (via a gamepad), indicating which interval contained the largest depth separation (depth magnitude variant) or which interval contained the most tangible, solid and real depth separation (depth realism variant). We exactly instructed our participants to "*judge in which of the two intervals the image had a bigger depth separation between the two planes*". The task used in the depth-realism variant was based on that used by Hibbard et al. (2017), and used terms derived from an analysis of the phenomenology of 3-D perception (Vishwanath & Hibbard, 2013). Each stimulus interval was presented for 3 sec with a 1 sec inter-stimulus interval (blank screen). Observers were instructed to look around the scene and fixate both stimulus rectangles, to ensure dynamic patterns of accommodation response and retinal blur were generated, as in real-world viewing. We exactly instructed our participants to "*judge in which of the two intervals the image had a more tangible, solid and real depth between the two planes*". We used time-sequential presentation rather than simultaneous presentation (c.f.Hibbard et al., 2017) because of the relatively restricted field-of-view of our display. Depth separations of 0.25, 0.375, 0.5, 0.625 and 0.75 D were presented in all focus-cue conditions. The conventional-stereo-3-D condition also included a depth separation of 0.14 D, which was not possible in the correct-focus-cues condition due to physical limitations of the display. The pairs of stimuli presented are shown in Figure 3.4. The minimum focal-plane spacing of the display meant that 'adjacent' pairs of depth separations (e.g. 0.375 vs. 0.5 D) could not be presented in a single trial in the correct-focus-cues condition. Otherwise, all possible combinations of depth separation and focus-cue condition were presented, within and across focus-cue conditions (a total of 116 comparisons; see Figure 3.4). Within each trial, the order that each pair of stimuli was presented in was chosen at random. Observers completed 20 repetitions of each stimulus pair, across six blocks of trials, completed across multiple days (2,320 trials in total for both depth-magnitude and depth-realism variants of the experiment). For the correct-focus-cues trials, a given positioning of the three focal planes allowed two depth separations per experiment block. Each block therefore contained two correct-focus-cues depth separations, and the remaining trials were chosen at random from all of the remaining pairwise comparisons.

## 3.3 Results

Each observer's data, for both their depth magnitude and depth realism judgments, were converted into Thurstone 'scale values' (Thurstone, 1927). In Thurstonian scaling, the scale values represent an underlying psychological continuum that cannot be directly observed or measured. Instead, they are inferred from the observed choices or rankings made by participants when presented with different pairs of stimuli (which could be images, sounds, or any other perceptual inputs). Thurstonian scaling is used when researchers want to understand how individuals perceive differences between stimuli rather than their absolute magnitudes or in the cases where absolute magnitudes of the stimuli cannot be practically observed. The scale values obtained indicate how strongly one stimulus is preferred or perceived to be different from another on the underlying psychological continuum. Participants are presented with pairs of stimuli and asked to make judgments about which one they perceive as stronger, larger, more realistic, or in some other way different. By analysing the participants' choices, researchers can rank the stimuli on the inferred psychological scale. The distance between two scale values indicates the perceived difference between the corresponding stimuli. A larger distance between two scale values suggests that the perceived difference is more substantial compared to a smaller distance. For each pair-wise comparison we calculated the proportion of times a given depth separation was chosen (collapsed across stimuli where the far rectangle was at the top and at the bottom), and used Bayesian maximum likelihood estimation under Thurstone's case V conditions to calculate scale values for each depth separation in each focus-cue condition (Tsukida & Gupta, 2011; Perez-Ortiz & Mantiuk, 2017).

We first examined each observer's data for evidence that they had responded preferentially to the rendered blur foil condition when making realism judgements.

Three observers showed a very strong preference for the foil condition, increasing with depth separation (and so the amount of rendered blur; Figure 3.5). Based on the logic outlined previously, we reasoned that these observers likely selected stimuli in which they perceived blur regardless of its effect on realism. Although there was no quantitative exclusion criterion, but visual inspection showed that the participants' responses were clearly bimodally distributed (see Appendix, Figure 1). Participants either showed a very clear preference for Gaussian blur (P4, P7 and P14) or no evidence of a preference at all (the remaining participants). There were no intermediate cases, showing slight preference for Gaussian blur. We therefore cannot be confident that any preference they showed for the correct-focus-cues stimuli over conventional stereo 3d presentation reflected actual increases in perceived realism (see right panel in Figure 3.5). We therefore removed these three observers from all subsequent analyses, including the analysis of depth magnitude scores.

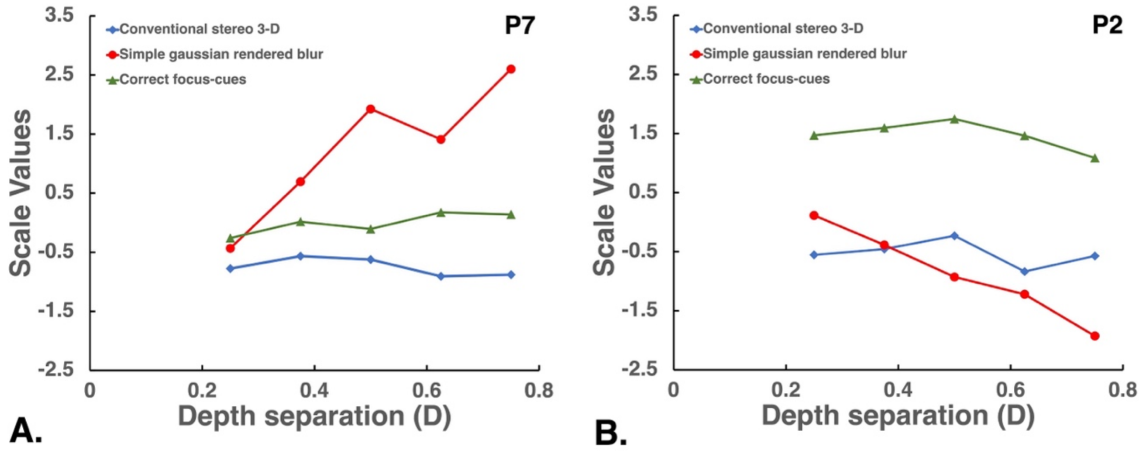



**Figure 3.5**. Example realism responses from individual observers. Depth-realism scores (derived via Thurstonian scaling) are plotted as a function of separation-in-depth of the pairs of stimulus planes for all the three focus cues conditions: conventional stereo 3d (blue), simple Gaussian rendered blur (red) and correct focus cues (green). **A**. A representative observer who did not consistently judge the foil condition as more realistic than the other conditions. **B**. A representative example of the three observers who showed a strong preference for the foil condition.

We averaged across the remaining 12 observers to give average scale values for depth magnitude and depth realism judgements. Figure 3.6 shows the results for both types of judgements, as a function of depth separation between the random-dot-defined rectangles, for the focus-cue conditions. Our analysis concentrates on the results of correct-focus-cues vs. conventional stereo presentation (Figure 3.6 also shows the results in the Gaussian blur foil conditions (red) for inspection purposes). Perceived depth magnitude (Figure 3.6 A) increased systematically as a function of increasing depth separation in both the conventional stereo and correct-focus-cues conditions. We did not find a roll-off in perceived depth at larger depth separations (c.f. Hibbard et al., 2017). There was no evidence of an effect of focus-cues condition on depth magnitude. Consistent with this, a two-way (depth separation x focus-cue condition) repeated measures ANOVA found a significant main effect of depth separation on perceived depth magnitude, ($F(4, 44) = 7.92$, $p < 0.00$), with a moderately significant effect size (partial eta squared of 0.418), but no significant main effect of correct-focus-cues vs. conventional stereo presentation, ($F(1, 11) = 0.49$, $p = 0.50$),with a minimal effect size (partial eta squared of 0.043), and no_significant interaction, (F(4, 44) = 1.52; p = 0.21), with a minimal effect size (partial eta squared of 0.121). These results indicate that while perceived depth increased reliably with disparity-specified depth separation, the magnitude of perceived depth was unaffected by focus-cue condition.

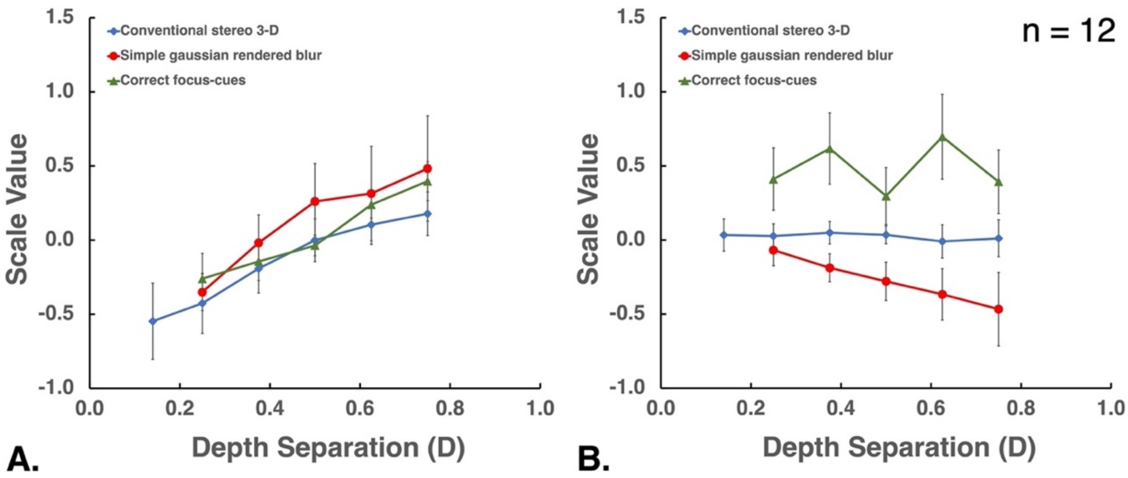



**Figure 3.6:** Experiment results. Depth-magnitude (left panel) and depth-realism scores (derived via Thurstonian scaling) are plotted as a function of separation-in-depth of the pairs of stimulus planes for all the three focus cues conditions (3d condition [Blue], rendered condition [Red] and natural condition [Green]). Error bars depict ±1 standard error of the mean. **A**. Depth magnitude judgement increases as depth separation increase in all conditions. **B**. Natural focus cues (green) were consistently judged as portraying more realistic depth separation than either conventional stereo 3d presentation (blue) or rendered Gaussian blur (red).

For depth realism judgements, correct focus cues were thematically judged as more realistic than conventional stereo presentation (Figure 3.6 B). This is consistent with correct focus cues providing a benefit to perceptual realism. Depth realism was unaffected by depth separation, however. This was the case not only in the correct-focus-cues condition (as predicted), but also in the conventional-stereo condition (i.e. we did not replicate the roll-off in depth realism reported by Hibbard et al., 2017, with conventional stereo presentation). A two-way (depth separation x focus-cues condition) ANOVA confirmed there was a significant main effect of correct focus cues vs. conventional stereo, (F(1, 11) = 5.08, p < 0.05), with a moderately significant effect size (partial eta squared of 0.316), no significant main effect of depth separation, ($F(4, 44) = 0.78$; $p = 0.55$), with a minimal effect size (partial eta squared of 0.066), and no significant focus cues x depth separation interaction, ($F(4, 44) = 1.16$; $p = 0.34$), with a minimal effect size (partial eta squared of 0.095). These data indicate that while perceived depth increased reliably with disparity-specified depth separation, the magnitude of perceived depth was unaffected by focus-cue condition.

## 3.4 Discussion

This experiment examined how presenting focus cues correctly, as opposed to conventional stereo 3-D presentation, affects comparative judgements of the magnitude of depth separation, and of the realism of depth separations (depth realism). A clear pattern of findings emerged. Correct focus cues did not alter the perceived magnitude of depth separation, but depth realism was consistently higher with correct focus cues compared to conventional stereo 3-D presentation. We explore these findings in more detail and discuss their implications for creating highly realistic stereo 3-D imagery, below.

*Depth magnitude judgements*

The results of the depth magnitude judgements demonstrate that perceived depth increased reliably in all of the focus-cues conditions with increasing portrayed depth. Primarily, this condition serves as an important control for interpreting the depth realism judgements. Identical stimuli were used in the depth magnitude and depth realism judgements, and the systematic pattern of depth magnitude judgements we observed confirms that observers could perceive depth in the stimuli they were asked to make realism judgements about.

We found that focus-cue condition did not affect the magnitude of perceived depth. As outlined in the Introduction, the pattern of effects in previous studies is inconsistent, and in our view this result is unsurprising. Previously, varying the stimulus to accommodation has been shown to affect the interpretation of disparities (Watt et al., 2005a), and so one might expect that the incorrect stimulus to accommodation at the far depth plane might result in compression of perceived depth compared to when focus cues were correct. Also, it has been shown that blur can be a relatively reliable cue to depth for objects at distances other than fixation (i.e. off the horopter, where disparity sensitivity is poorer; Held et al., 2012)—an analogous situation to fixating one stimulus plane our study and judging the depth of the non-fixated plane. Our observers looked between the two depth planes, sequentially, however, and so vergence may have provided a reliable and accurate signal to the relative distances of the planes even in the incorrect focus-cues conditions. Moreover, other studies have shown that perception of disparity-defined surface slant is unaffected by focus cues (Watt et al., 2005a,b).

In our study, perceived depth magnitude increased linearly throughout the range of separations tested, in contrast to Hibbard et al.'s (2017) finding of a 'roll-off' in depth magnitude at around 0.4 D depth separation. Hibbard et al. (2017) concluded that this was due to the stimuli exceeding binocular fusional limits, which would also be exceeded in our case. Our stimuli and task differed from those used by Hibbard et al. (2012), however, which may explain why depth magnitude continued to increase in our study. First, our depth planes were non-overlapping, rather than 'transparent'. Stereo transparency presents a greater challenge for disparity processing due to the prevalence of 'false matches' in the two eye's images (where stereoscopic matching can occur between non-corresponding points in the two eye's images; Howard and Rogers, 2002). This means that a clear percept of stereoscopic depth can break down relatively easily for stereo-transparency stimuli. Second, as noted above, our observers were explicitly instructed to look between the two depth planes, and so made sequential vergence eye movements, which could have resulted in accurate perception of depth magnitude even when the non-fixated stimulus was unfused, similar to the phenomenon of ‘sequential stereopsis’ (Enright, 1991).

*Correct focus cues increase depth realism*

Our results for depth realism judgements showed that correct focus cues resulted in higher depth realism scores than with conventional stereo 3-D presentation. As noted by Hibbard et al. (2017), the fact that the pattern of effects of both focus-cue condition and depth separation were very different across depth magnitude and depth realism judgements also suggests that observers were able to respond selectively to different dimensions of their perceptual experience in the two tasks. Taken together, these findings indicate that presenting correct focus cues does increase depth realism, and that presenting focus cues correctly (or at least approximating them; see below) may therefore be necessary if realism in stereo 3-D is to be maximised.

Consistent with our expectations, when focus cues were presented correctly, depth realism did not decrease with increasing depth separation, but instead remained on average constant. This is in keeping with the intuition that the realism of real-world scenes is not normally affected by depth and distance relations. We also did not find a roll-off in depth realism with depth separation in our conventional stereo 3-D condition, however, which Hibbard et al. (2017) reported (they used conventional stereo 3-D presentation). We speculate that this difference, too, may be due to Hibbard et al.’s (2017) use of transparent stimuli. Difficulty fusing such stimuli (see above) results in a particularly confusing appearance, that may be unlike real-world experience. Further experiments specifically comparing transparent to non-overlapping stimuli would be required to determine whether this is the case.

It is interesting to note that correct focus cues confer a similar increase in depth realism independent of depth separation. This is perhaps surprising, given that the magnitude of error in focus cues in the conventional stereo 3-D condition increases directly with increasing depth separation. Our data suggest that detectable presence of incorrect focus cues determines the effect on depth realism, and not the precise magnitude of the error.

In a broad sense, Hibbard et al. (2017)'s results might be thought of as revealing properties of perception in conventional 3-D displays, whereas the current study (correct focus cues condition) investigates properties of perception in conditions closer to natural viewing. This potentially allows more general conclusions to be drawn, and in particular that focus cues are likely to contribute to the subjective sense of three dimensionality of scene structure in real-world viewing.

We note that correct focus cues resulted in increased depth realism even at the smallest depth separation we tested of 0.25 D. This may seem surprising at face value, because the retinal images for the correct-focus-cues and conventional-stereo-3-D stimuli would be quite similar. In fact there are reasons to expect that these conditions would be discriminable, however. First 0.25 D is close to the effective depth of focus of the eye (estimated to be ~0.25 to 0.3 D; Campbell, 1957; Charman & Whitefoot, 1977; Rolland et al., 1999) and so the retinal blur of the non-fixated plane in the correct-focus-cues condition may be detectable even in static viewing. Indeed, the likelihood of this is increased by the fact our stimuli had low luminance, and were sparse, and so the pupil size would have been large, reducing the eye's depth of focus. Second, additional signals beyond first-order retinal blur likely play a role in discriminating the two types of stimuli. Chromatic aberrations and higher-order aberrations both indicate the sign of a separation in depth (Fincham, 1951; Kruger et al., 1993; Aggarwala et al., 1995; Lee et al., 1999; Fernandez & Artal, 2005; Chen et al., 2006; Chin et al., 2009a; Cholewiak et al., 2017), and the presence of these correct signals in the correct-focus-cues condition may signal increased depth realism even for small depth separations. Also, the eye makes continuous microfluctuations in accommodation, resulting in patterns of contrast change that depend on the relative depths of objects (Campbell et al., 1959; Charman & Tucker, 1978; Kotulak & Schor, 1986; MacKenzie et al., 2010). Consider fixating the near stimulus plane, for example. Here, with correct focus cues, outward microfluctuations of accommodation will result in increased retinal contrast at the far plane, and reduced retinal contrast at the near plane. In conventional stereo 3-D presentation the same change in accommodation results in perfectly correlated contrast changes in the two stimulus planes. Although these signals are very fine-scale, there is evidence that they drive the accommodation response effectively, even at small defocus errors (Fincham, 1951; Campbell et al., 1959; Charman & Tucker, 1978; Kotulak & Schor, 1986; Kruger et al., 1993; Aggarwala et al., 1995; Lee et al., 1999; Fernandez & Artal, 2005; Chen et al., 2006; Chin et al., 2009a; MacKenzie et al., 2010; Cholewiak et al., 2017). They could also therefore create perceptible changes in realism. Finally, even at 0.25 D depth separation, the correct focus cues condition likely stimulates an accommodation response, which would both produce changing patterns of retinal contrast similar to microfluctuations, above, and the motor response itself—which is not present in conventional stereo 3-D viewing—may also be detectable.

*Challenges in assessing subjective aspects of depth realism*

It is not universally accepted that qualitative aspects of depth perception, and stereopsis, can be meaningfully separated out from quantitative aspects (e.g. Rogers, 2019). Clearly, depth realism judgements are more subjective than depth discrimination judgements, and the factors contributing to them are difficult to identify, evaluate, and control. It is therefore difficult to be certain that our observers judged depth realism on the basis we intended (or even whether this is possible). To minimise variability in how the task was completed, we gave carefully worded instructions, and used a pairwise comparison approach (and Thurstonian scaling) that is agnostic with respect to how observers are making their distinction. Ultimately, however, in our conception, the aim of 'realistic computer graphics'—to create realistic-looking imagery—is inherently subjective, and so can only be assessed by probing subjective aspects of perceptual experience.

*Limitations and future work*

While our findings indicate that correct focus cues can increase depth realism, the experiment was intentionally designed to isolate the role of focus cues in stereo 3-D. As such, it provides proof-of-principle evidence that focus cues play a role in perceptual realism, but it does not answer the in-practice question of whether correct focus cues result in increased perceptual realism in typical stereo 3-D use cases where 'hyper-realism' might be important. We discuss several reasons for this below.

We presented objects in a reduced-cue environment, where depth was specified only by focus cues and binocular disparity. As noted above, this allowed us to isolate the contribution of focus cues in the absence of other uncontrolled variables. Real-world applications for highly realistic stereo 3-D imagery will include the normal range of depth cues present in the real world, however. It is not clear whether in such scenes correct focus cues would be expected to be less important for perceptual realism, or whether they will still play a clear role in increasing realism. Viewed in the framework of depth-cue integration (e.g. Knill and Saunders, 2003; Hillis et al., 2004), adding more cues to 3-D structure (perspective, shading etc.) would be expected to cause a down-weighting of focus cues, leading to them having a smaller influence on perception. This theory applies to how different cues to the *magnitude* of perceived depth are combined, however, and not realism per se (and note, focus cues did not affect depth magnitude in our current reduced-cue situation). In natural viewing changes in retinal blur affect the appearance of scene as a whole, in a way that is predictable given knowledge of 3-D scene structure. So, for example, all the details of an object farther than fixation are blurred on the retina and become sharp when we move fixation to it and accommodate appropriately. The presence of multiple depth cues will not affect the visibility of these changes in retinal blur (they may even be more noticeable for finely detailed real-world objects). Nor will motor signals form the accommodation response differ from those in reduced-cue scenes. So, it is entirely possible that correct focus cues will have similar, positive effects on depth realism in highly realistic stereo 3-D scenes. We address this question in Chapter 4.

The 3-D structure of our scenes was highly constrained, compared to natural scenes. We used isolated, planar objects because that allowed us to present truly correct focus cues on our multiple-focal-planes display. This was important because our intention was to answer the fundamental question of whether correct focus cues are important for depth realism, and not to evaluate the efficacy of our display approach per se. While interpolation techniques (so-called depth-filtering, or depth-weighted blending, where image intensity is distributed across focal planes) can stimulate the accommodation response correctly to intermediate distances between planes (MacKenzie et al., 2010), the pattern of retinal blur is nonetheless significantly different than that derived from viewing equivalent real scenes (Akeley et al., 2004; MacKenzie et al., 2010; Narain et al., 2015). If we presented more complex scenes—objects that contained depth relief, and supported by a continuous ground plane, for example—we would be knowingly presenting at least partially incorrect focus cues, conflating the role of focus cues with the efficacy of the display approach. This issue applies to all existing efforts at presenting focus correctly, which necessarily involve some degree of inaccuracy or trade-off (Pastoor & Wöpking, 1997). Much as with adding further cues to depth, it is not clear to us a priori whether more complex scene 3-D structure would be expected to influence the contribution of correct focus cues to depth realism. More information from focus cues is likely to be present due to more complex scene structure. However, more continuous variation in depth (for example along a ground plane) could conceivably result in less salient blur changes compared to the step changes in our current stimuli. Further research is needed to disentangle these possibilities, but presenting correct focus cues for such scenes currently exceeds technological capabilities.

Many real-world scenes have much higher luminance than our experimental stimuli, and so the pupil size will be smaller. As discussed previously, a smaller pupil increases the depth of focus of the eye (i.e. a given defocus error results in less retinal blur), and this in turn increases tolerance to accommodation errors. In principle this would reduce the differences in retinal blur and accommodation response between stimuli presented with conventional stereo 3d and with correct focus cues. At least for moderate depth separations between objects, this could result in smaller effects of correct focus cues on depth realism compared to our reduced-cue study.

Finally, because our experiment compared fully correct focus cues to conventional stereo 3-D presentation, we cannot determine the contribution of the different signals from correct focus cues to the increase in depth realism. For example, we cannot know whether changes in retinal blur, the presence of a motor response, or both are contributing significantly to the increased realism. It is of interest to understand this because various approaches have been proposed for addressing incorrect focus cues, and which aim to present focus cues with differing degrees of correctness ranging from approximating only relative blur to attempts at reproducing the natural light field at the eye. The complexity (and cost) of the solutions also tends to increase with increasing fidelity of focus-cue presentation. The extent to which different aspects of focus cues contribute to realism could therefore be useful in determining which approaches to presenting focus cues should be adopted.

Arguably the most conceptually straightforward approach is gaze-contingent rendering. Here, eye-tracking is used to determine the current fixation point, and other points in the scene are rendered with artificial blur in accordance with their focal distance relative to fixation (e.g. Duchowski et al., 2014; Vinnikov & Allison, 2014; Maiello et al., 2014). The accuracy of the resulting blur depends on the accuracy of the blur model used, including whether it incorporates individualised eye-optics, chromatic aberration etc. (Cholewiak et al., 2017). Yet in this approach, the display remains at a fixed focal distance. The stimulus to accommodation is therefore incorrect because it does not vary with 3-D scene structure. Similarly, accommodation microfluctuations do not result in correct changes to retinal contrast. A more complex variant combines gaze-contingent blur rendering with a variable focus element, so that the display plane is moved to the optical distance of the currently fixated scene point (Hasnain et al., 2019). Such displays in principle stimulate accommodation near-correctly, though the dynamic changes in blur may not match the natural changes during an accommodation response. And because the scene is at a constant focal distance during fixation, microfluctuations in accommodation will again produce incorrect patterns of retinal contrast variation. As discussed, multiple-focal-planes displays (more correctly, a type of fixed-viewpoint volumetric display; Akeley et al., 2004) do not rely on gaze tracking, but rely on interpolation which can approximate the natural stimulus to accommodation, but not fine-scale aspects of retinal blur at intermediate distances between planes, and cannot simulate occlusions correctly (Akeley et al., 2004; MacKenzie et al., 2010; Narain et al., 2015). Light-field displays, and holography-based approaches hold promise of a closer approximation to the natural light field, but are not yet practical, and still make compromises on other aspects of image quality that may degrade realism. It therefore remains to be determined what the best approach is to present correct focus with a view to maximising perceptual realism.

# Chapter 5

## 5 General Discussion

### 5.1 Overview

Creating realistic stereo 3-D scenes is a very active field of research right now, and advances in graphic hardware, rendering techniques, and display has made presenting most cues rather accurately possible (de Silva et al., 2011; Banks et al., 2016; Zhong et al., 2021). But focus cues is hard to get right in conventional stereoscopic 3-D displays (Akeley et al., 2004). Focus cues are depth cues (Mather, 1997; Buckley & Frisby, 1993; Watt et al., 2005a; Hoffman et al., 2008), and upon getting it wrong in a stereoscopic 3-D display, ensues visual discomfort, decreased stereo depth perception, slower stereo fusion, etc (Wann & Mon-Williams, 2002; Niida & Okano, 2005; Watt et al., 2005a, 2005b; Hoffman et al., 2008; Lambooij et al., 2009). And these effects give a signal to un-naturalness of the depicted scene, because, when viewing in real world we do not encounter any issues arising from incorrect focus cues. So, focus cues should be responsible for things to appear realistic (Hibbard et al., 2017). There are several potential display technologies that can deliver correct focus cues (Favolora et al., 2002; Lucente et al., 1997; McQuaide et al., 2002; Akeley et al., 2004; Chang et al., 2018; Zhong et al., 2021). However, these are not commercially viable options, as they are rather expensive, complex, and/or gets focus cues right on the expense of important aspects of vision (image resolution, transparency and specular highlights in the scene, and dynamic range). So, the question that we are asking in this thesis is, if it is worth getting focus cues right for producing realistic scenes, and if we could optimise the stereo 3-D content so that the effect of incorrect focus cues can be mitigated in conventional displays. And if 3-D content optimization is not possible for overcoming issues arising from incorrect focus cues, then we will have no other choice but research on technologies that can get focus cues correct without compromising other depth cues. In this general discussion we will report the results from our empirical chapter 2 to 4.

## 5.2 Optimization of 3-D content for mitigating vergence-accommodation conflict

In chapter 2, we tried to provide guidelines on optimising stereo 3-D content across wide ranges of audiences, by simple using basic demographic information such as their age. This was necessary because, getting forces cues correct is difficult as it involves development of complex hardware, and it will be costly to replace existing display technologies esp. conventional 3-D stereo displays. So, if focus cues are to have significant contribution to perceived realism, then optimising existing display and 3-D content would enable to keep using existing displays. And if we can find how the visual factors such as age, ability to accommodate, phoria, and /or ZCSBV effects the tolerance to conflicts, we could adapt the content to suit the audiences or, we could adapt audience's zone of tolerance to the 3-D content using prisms and lenses.

However, we did not find any relationship in peoples zone of good stereo depth perception with age and their ability accommodate, but phoria seems to predict the zone centre positions of individuals, similar to what Shibata found (Shibata et al., 2011), but this is not very practical measure as most people do not know their phoria and the relationship between phoria and the zone centre is very noisy. So, cannot be used as a guideline to go ahead and tailor 3-D content.

*Zone of 'comfort' and 'good stereo'*

In Shibata's study, they have underlined the possibility of constructing 3-D content within a depth budget where people are reasonable tolerant to visual discomfort caused by vergence-accommodation conflict. They found for a fixed level of disparity, increased viewing distance gives a progressively larger 'zone of comfort'. So, for a given range of disparity people are more comfortable when they view the content from a farther distance. Making 3-D cinema the best option for the application of stereo 3-D, and head mounted display the worst for viewing large disparities without visual discomfort.

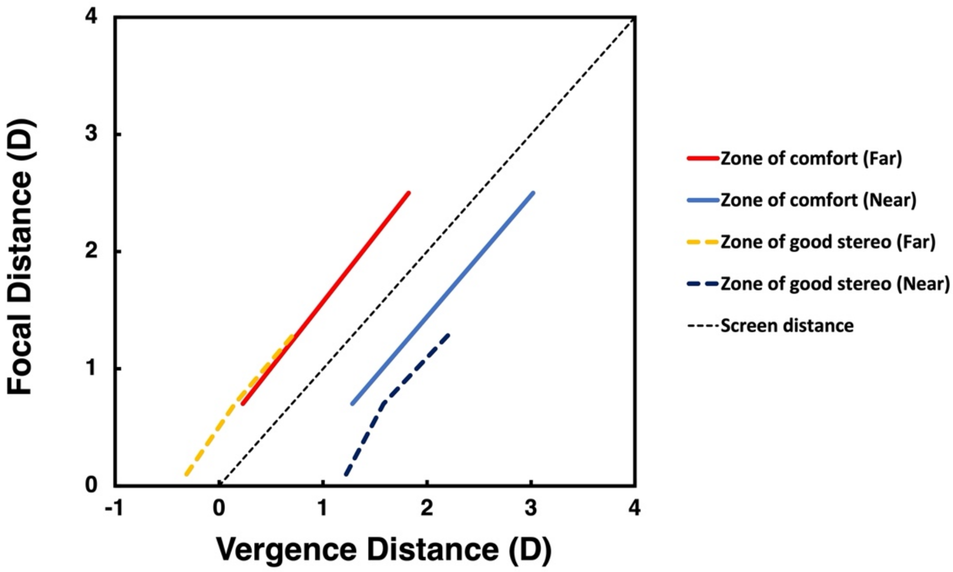



**Figure 5.1**: A comparison plot between 'Zone of comfort' and 'Zone of good stereo'. The far and near boundaries of both 'Zone of comfort' and 'Zone of good stereo' are plotted as a function of vergence and focal distances in dioptres. The far boundaries for 'Zone of comfort' [solid red line] and 'Zone of good stereo' [dashed yellow line] corresponds to the estimates of the largest comfortable negative conflicts (content behind the screen) and the near boundaries for 'Zone of comfort' [solid light blue line] and 'Zone of good stereo' [dashed dark blue line] corresponds to the estimates of the largest positive conflict (content in front of the screen). The dashed black line represents the screen distance.

And this is very useful especially for adapting existing technologies to match viewer's tolerance to vergence-accommodation induced visual discomfort.
We believe that our data too can provide at least ranges of conflict distances that viewer's might be tolerant to degradation of stereo depth perception and help content creators produce stereo 3-D content that is suitable to a wide range of audiences.
The pathway of how vergence-accommodation conflict effects both visual comfort and stereo depth perception are closely linked, so, we speculate that, the sizes of conflict that causes visual discomfort and degradation in stereo depth should match/overlap with eachother, at least in theory.

To see if the zone of comfort overlaps with the zone of good stereo depth perception, we plotted the width (boundaries) of 'zone of comfort' and 'zone of good stereo' as a function of vergence and focal. We derived estimates for far and near boundaries of the zone of comfort from equation 7 in Shibata's paper (Shibata et al., 2011). Our screen distances (1.3D, 0.7D, and 0.1 D) were relatively small compared to Shibata's (2.5D, 1.3D, and 0.1D). We see that both zone of good stereo and zone of comfort are quite comparable. The farther boundaries of both zones overlap somewhat, however the near boundaries do not (Figure 5.1). Anyway, we would like to acknowledge the fact that these zone ('comfort' and 'good stereo') data cannot be compared directly, and is a reasonable speculation at best.

*Range of vergence distances available for content creators*

We can derive range of vergence distances around the screen for producing a 3-D content which does not induce visual discomfort and degradation in stereo depth perception by confining the depth budget in the zone of 'comfort' and 'good stereo depth perception'. According to 'zone of comfort', for a screen distance of 1.3 D the nearest conflict (from the screen) that can be presented is 1.86 D and farthest is 0.76D, and for a screen distance of 0.7 D the nearest conflict (from the screen) that can be presented is 0.7 D. And according to the 'zone of good stereo', for a screen distance of 1.3 D the nearest conflict (from the screen) that can be presented is 0.91 D and the farthest is -0.57 D, and for a screen distance of 0.7D the nearest conflict (from the screen) that can be present is 0.88 and farthest is -0.54D. These depth ranges are quite large for stereoscopic 3-D content to be presented without inducing any visual discomfort and any degradation in stereo depth perception.

## 5.3 Implication of focus cues on 'depth realism'

In chapter 3 and 4, we looked at how incorrect presentation of focus cues effects depth realism in a scene. In chapter 3, we compared stereoscopic images with correct focus cues to conventional stereoscopic images with incorrect focus cues for a range of depth separations. We found significant effect of focus cues on depth realism. Interestingly participants were able to differentiate stimulus with correct focus cues being more realistic than conventional stereo 3-D (incorrect focus cues) for smallest depth separations of 0.25D. Implying that getting focus cues correct is necessary when creating display solutions intended to produce highly realistic 3-D contents. We were able to replicate similar results in chapter 4, in a completely different stereoscopic 3-D display which worked on similar principles to ours, with high fidelity stereoscopic images. This was important as, our visual system does not rely on only one depth cue, but it derives depth information from a variety of depth cues (Backus & Banks, 1999; Buckley & Frisby, 1993; Jacobs, 1999; Körding & Wolpert, 2004) and integrates them in optimal or near optimal fashion to produce depth estimates with minimum-variance (Jacobs, 1999; Knill & Saunders, 2003; Hillis et al., 2004). And there is evidence supporting that when combining sensory signals, the reliability of the cue is considered (Backus & Banks, 1999; Buckley & frisby, 1993; Jacobs, 1999; Körding & Wolpert, 2004; van Beers et al., 1998). So, it is possible that our visual system may totally ignore depth estimates from focus cues, or it might get masked by other less noisy depth cues in a near realistic scene where other depth cues are fairly accurate. We were able to establish that, correct focus cues had a significant effect on realism even in a depth cue rich environment.

*Implications for approaches to presenting correct focus cues*

For presenting focus cues correctly both stimulus to accommodation and retinal blur gradient should be correct. But most of the conventional displays only get either one of them correct. And this cause vergence-accommodation conflict resulting in visual discomfort and decreased stereo depth perception and other related issues. But it should certainly reduce the perceived realism of the scene. Studies have tried to come up with display solutions which try to present focus cues correct (Favolora et al., 2002; Lucente, 1997; McQuaide, 2002; Schowengerdt & Seibel, 2006; Sullivan, 2004; Akeley, et al., 2004; MacKenzie et al., 2010; Chang et al., 2018; Zhong et al., 2021).

Gaze-contingent displays were developed for the purpose of presenting correct stimulus to accommodation. These displays use an eye tracker to track viewer's gaze and perform a blurring of the region in the scenes other than where the viewer is fixating to (Duchowski et al., 2004; Biebl et al., 2022). This does not help in driving accommodation correctly as the light rays are still coming from the display surface. It is difficult to track the eyes accurately, and apply the required blurring appropriately (Biebl et al., 2022). Changes in vergence angles once you get to large distances are very minute and are virtually un-trackable. Eye trackers are not very sensitive to small deviating in vergence, and since you can have very large change in depth gradient for very small change in visual angles. And it is almost impossible to correctly simulate transparent surfaces in the 3d scene, and this cannot be solved with gaze contingent blurring. And gaze-contingent displays do not get retinal blur correct either. Because even if correct retinal blur is rendered, the visual system cannot make out the sign of the blur. It needs associated cues such as chromatic aberration in retinal image to prove sign to the retinal blur gradient (Nguyen, et al., 2005; Cholewiak et al. 2017). But there are quite a few gaze-contingent displays available and in use (Reingold et al., 2003; Loschky & Wolverton, 2007). These gaze-contingent display are quite expensive and to our understanding, appears to be quite ineffective at what it promises to deliver. Some studies have found gaze-contingent displays have reduces visual discomfort when viewing stereo 3-D content (Duchowski et al., 2014) and improve stereoscopic vision (Maiello et al., 2014). However, this might not hold for bigger conflict sizes. And research have found that people dislike resultant images because of temporal lag in real-time data transmission between the gaze-contingent and software components (Wann et al., 1995; Vinnikov and Allison, 2014).

It is possible to get stimulus to accommodation correct in a gaze contingent display, but only if it incorporates a varifocal system, where the entire display moves in space (across the focal depth) to present correct stimulus to accommodation and the image is generated in time multiplexed fashion (Favalora et al., 2002; Schowengerdt & Seible, 2006). However, these displays still do not get stimulus to retinal blur correct because of fore mentioned reasons. This type of display would be ideal for scenarios where presenting depth realism is not a priority. But rather useful for presenting stereo 3-D content that does not create vergence-accommodation conflicts. Thought this may seem attractive option but the display system will be quite expensive and complex, at it would however more useful to create a fixed view-point volumetric (multiple-focal plane) display as it could both aspects of focus cues (accommodation & retinal blur) accurately.

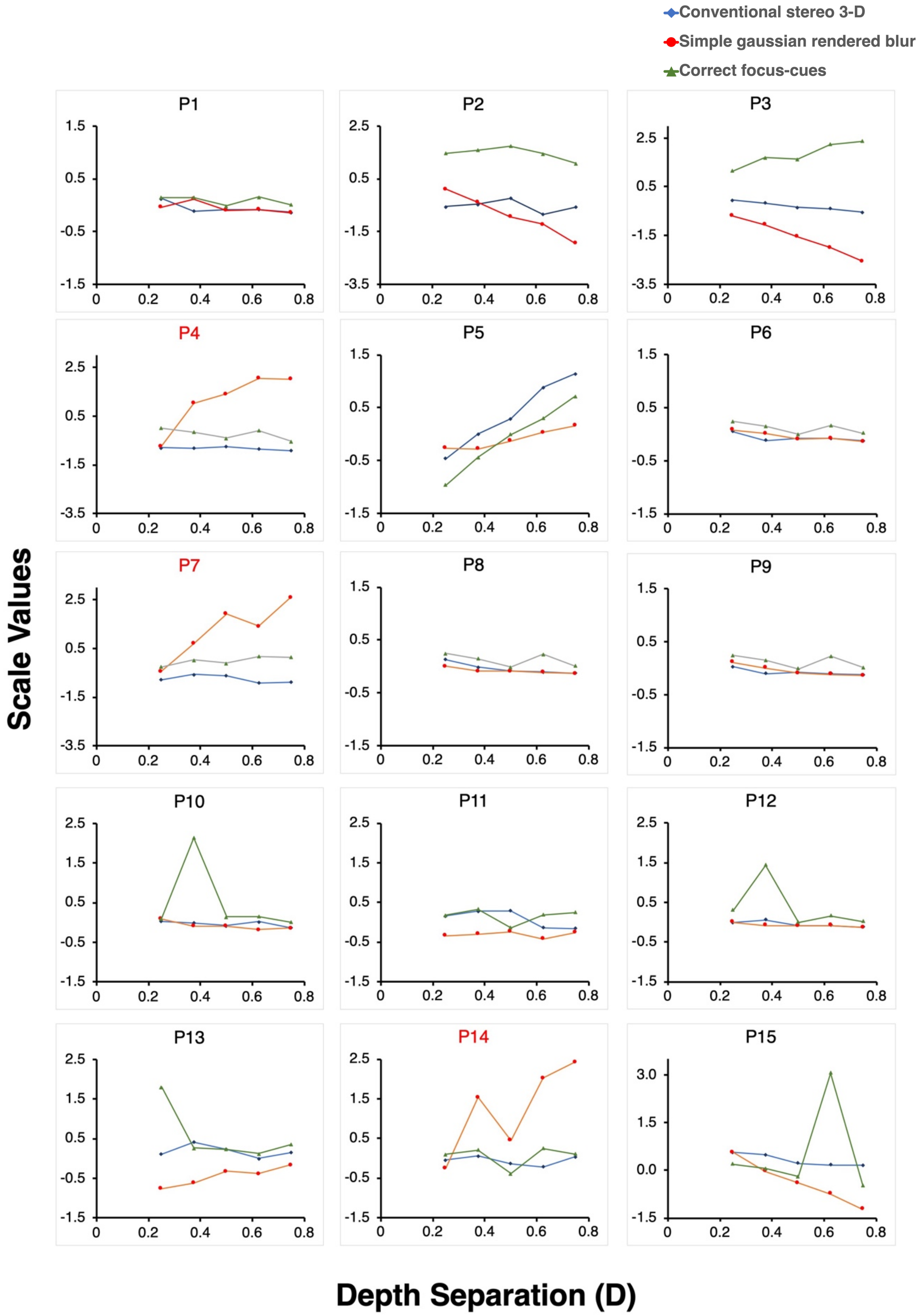
A fixed-viewpoint volumetric display may not support multiple viewers and multiple viewpoint capabilities, but it does provide important advantages. In a fixed-view point display as the name suggests, the view dependent lighting effects can be computed correctly for each eye's single line of sight, because the viewer's position is fixed. And fixing the viewpoint also eliminates the requirement for equal resolution in all dimensions of the scene. The visual system's resolution of focal depth is way poorer than spatial resolution (Charman & Whitefoot, 1997; Campbell, 1957). So, we can achieve sufficient spatial resolution and focal-depth resolution with few focal planes (Akeley et al., 2004). And then using optical elements such as a Badal lens (positive spherical lens) large focal distances up to optical infinity (MacKenzie et al.,2010). There are many different types of fixed-viewpoint volumetric displays, but they are all based on the same concept. Most prominent and widely used fixed-viewpoint volumetric display is a multiple-focal planes display, and we have used this display type in all of our experiments (details of our display has been discussed in [section 1.7]. And from our study's perspective, we have found fixed-viewpoint volumetric displays quite capable at presenting both aspects of focus cues correct. And hence has the potential to provide conflict free and realistic looking stereoscopic 3-D imagery.

## 5.4 Conclusion

We can show sufficiently large ranges of stereoscopic depth on a conventional display without encountering degradation in stereo depth perception due to vergence-accommodation conflict. However, there is a cost to realism if we get focus cues incorrect. Human visual system seems to be very sensitive to inaccuracies in stimulus to focus cues when judging the realism of depth in a scene. So, getting focus cues correct is crucial when building display that seeks to present highly realistic scenes. And conventional display solutions such as gaze-contingent and varifocal displays do not provide appropriate stimulus to correct focus cues. So, a decision needs to be made based on the particular requirements of the application.

Display that falls in the class of fixed-viewpoint volumetric displays seems to provide optimal solutions to both realism and vergence-accommodation conflict free imagery. However, we need to acknowledge the fact that getting focus cues correct is technological and financially difficult, at least with current available technologies. And fixed-viewpoint volumetric displays do not provide a practical solution for commercial applications. Also, the cost of replacing current display technology (cinema screens, 3-D televisions, and head mounted displays (HMD's)) would be very high and impractical. So, the best solution would be to optimize the content across individual users to mitigate the effects of vergence-accommodation conflict. But, if the agenda is to create highly realistic stereo imagery, then more research needs to be done pertaining to miniaturization of existing multi-focal planes display and their elements, improving graphical computation methods and associated hardware.

# 6 Appendix




**Appendix Figure 1**: Realism judgement of all observers. Depth-realism scale values (derived via Thurstonian scaling) are plotted as a function of separation-in-depth of the pairs of stimulus planes for all the three focus cues conditions (Conventional stereo 3-D [Blue], Simple gaussian rendered blur condition [Red] and Correct focus-cues condition [Green]).

# 7 References


Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology : CB*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Aggarwala, K. R., Kruger, E. S., Mathews, S., & Kruger, P. B. (1995). Spectral bandwidth and ocular accommodation. *J. Opt. Soc. Am. A*, *12*(3), 450–455. https://doi.org/10.1364/JOSAA.12.000450

Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A Stereo Display Prototype with Multiple Focal Distances. *ACM Trans. Graph.*, *23*(3), 804–813. https://doi.org/10.1145/1015706.1015804

Banks, M. S., Hoffman, D. M., Kim, J., & Wetzstein, G. (2016). 3D Displays. *Annual Review of Vision Science*, *2*(1), 397–435. https://doi.org/10.1146/annurev-vision-082114-035800

Backus, B. T., & Banks, M. S. (1999). Estimator Reliability and Distance Scaling in Stereoscopic Slant Perception. *Perception*, *28*(2), 217–242. https://doi.org/10.1068/p2753

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, *33*(7), 919–933. https://doi.org/10.1016/0042-6989(93)90075-8

Borel, T., & Doyen, D. (2013). 3D Display Technologies. In *Emerging Technologies for 3D Video* (pp. 295–312). https://doi.org/10.1002/9781118583593.ch15

Brewer-Deluce, D., Bak, A. B., Simms, A. J., Sinha, S., Mitchell, J. P., Shin, D., Saraco, A. N., & Wainman, B. C. (2021). Virtual Reality Bell-Ringer: The Development and Testing of a Stereoscopic Application for Human Gross

Anatomy. *Anatomical Sciences Education*, *14*(3), 330–341. https://doi.org/10.1002/ase.2074

Cakmakci, O., & Rolland, J. (2006). Head-Worn Displays: A Review. *Display Technology, Journal Of*, *2*, 199–216. https://doi.org/10.1109/JDT.2006.879846

Campbell, F. W. (1957). The Depth of Field of the Human Eye. *Optica Acta: International Journal of Optics*, *4*(4), 157–164. https://doi.org/10.1080/713826091

Campbell, D. T., & Fiske, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. Psychological Bulletin, 56, 81-105. http://dx.doi.org/10.1037/h0046016

Cochran, W. G. (1937). Problems Arising in the Analysis of a Series of Similar Experiments. *Supplement to the Journal of the Royal Statistical Society*, *4*(1), 102–118. https://doi.org/10.2307/2984123

Charman, W. N., & Tucker, J. (1977). Dependence of accommodation response on the spatial frequency spectrum of the observed object. *Vision Research*, *17*(1), 129–139. https://doi.org/10.1016/0042-6989(77)90211-5

Charman, W. N., & Whitefoot, H. (1977). Pupil Diameter and the Depth-of-field of the Human Eye as Measured by Laser Speckle. *Optica Acta: International Journal of Optics*, *24*(12), 1211–1216. https://doi.org/10.1080/713819479

Chin, S. S., Hampson, K. M., & Mallen, E. (2009). Role of ocular aberrations in dynamic accommodation control. *Clinical & experimental optometry*, *92*(3), 227–237. https://doi.org/10.1111/j.1444-0938.2009.00361.x

Cholewiak, S. A., Love, G. D., Srinivasan, P. P., Ng, R., & Banks, M. S. (2017). Chromablur: Rendering Chromatic Eye Aberration Improves Accommodation and Realism. *ACM Trans. Graph.*, *36*(6). https://doi.org/10.1145/3130800.3130815

Cumming, B. G., & Judge, S. J. (1986). Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey. *Journal of neurophysiology*, *55*(5), 896–914. https://doi.org/10.1152/jn.1986.55.5.896

Duchowski et al. (2014), Proceedings of the ACM Symposium on Applied Perception 2014, 39-46

De Silva, V., Fernando, A., Worrall, S., Kodikara Arachchi, H., & Kondoz, A. (2011). Sensitivity Analysis of the Human Visual System for Depth Cues in Stereoscopic 3-D Displays. *IEEE Transactions on Multimedia*, *13*(3), 498–506. https://doi.org/10.1109/TMM.2011.2129500

Emoto, M., Niida, T., & Okano, F. (2005). Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. *Journal of Display Technology*, *1*(2), 328–340. https://doi.org/10.1109/JDT.2005.858938

Enright J. T. (1991). Exploring the third dimension with eye movements: better than stereopsis. *Vision research*, *31*(9), 1549–1562. https://doi.org/10.1016/0042-6989(91)90132-o

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Fernández, E. J., & Artal, P. (2005). Study on the effects of monochromatic aberrations in the accommodation response by using adaptive optics. *J. Opt. Soc. Am. A*, *22*(9), 1732–1738. https://doi.org/10.1364/JOSAA.22.001732

Fincham, E. F. (1951). The accommodation reflex and its stimulus. *The British journal of ophthalmology*, *35*(7), 381–393. https://doi.org/10.1136/bjo.35.7.381

Fincham, E. F., & Walton, J. (1957). The reciprocal actions of accommodation and convergence. *The Journal of physiology*, *137*(3), 488–508. https://doi.org/10.1113/jphysiol.1957.sp005829

Fisher, S. K., & Ciuffreda, K. J. (1988). Accommodation and apparent distance. Perception, 17, 609-621.

Formankiewicz, M., & Mollon, J. (2009). The psychophysics of detecting binocular discrepancies of luminance. *Vision Research*, *49*, 1929–1938. https://doi.org/10.1016/j.visres.2009.05.001

Frisby, J. P., Buckley, D., & Horsman, J. M. (1995). Integration of Stereo, Texture, and Outline Cues during Pinhole Viewing of Real Ridge-Shaped Objects and Stereograms of Ridges. *Perception*, *24*(2), 181–198. https://doi.org/10.1068/p240181

Fry, G. A. (1939). Further experiments on the accommodation-convergence relationship*. *Optometry and Vision Science*, *16*(9).

Gao, R., Fan, X., Zhang, J., & Luo, Z. (2012). Haze filtering with aerial perspective. 2012 19th IEEE International Conference on Image Processing, 989–992. https://doi.org/10.1109/ICIP.2012.6467028

Gepshtein, S., & Banks, M. S. (2003). Viewing Geometry Determines How Vision and Haptics Combine in Size Perception. *Current Biology*, *13*(6), 483–488. https://doi.org/10.1016/S0960-9822(03)00133-7

Ghahramani, Z., Wolptrt, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In P. Morasso & V. Sanguineti (Eds.), *Advances in*

*Psychology* (Vol. 119, pp. 117–147). North-Holland. https://doi.org/10.1016/S0166-4115(97)80006-4

Gombrich, E. H. (1969). The evidence of images. In C. S. Singleton (Ed.), Interpretation: Theory and Practice (pp. 35--104). Johns Hopkins University Press.

Hasnain, A., Laffont, P.-Y., Jalil, S. B. A., Buyukburc, K., Guillemet, P.-Y., Wirajaya, S., Khoo, L., Deng, T., & Bazin, J.-C. (2019). Piezo-actuated varifocal head-mounted displays for virtual and augmented reality. In J.-H. Lee, Q.-H. Wang, & T.-H. Yoon (Eds.), *Advances in Display Technologies IX* (Vol. 10942, p. 1094207). SPIE. https://doi.org/10.1117/12.2509143

Henn, J. S., Lemole, G. M., Jr, Ferreira, M. A., Gonzalez, L. F., Schornak, M., Preul, M. C., & Spetzler, R. (2002). Interactive stereoscopic virtual reality: a new tool for neurosurgical education. Technical note. *Journal of neurosurgery*, *96*(1), 144–149. https://doi.org/10.3171/jns.2002.96.1.0144

Held RT, Cooper EA, O'Brien JF, Banks MS. Using Blur to Affect Perceived Distance and Size. ACM Trans Graph. 2010 Mar 1;29(2):19. doi: 10.1145/1731047.1731057. PMID: 21552429; PMCID: PMC3088122.

Hibbard, P. B., Haines, A. E., & Hornsey, R. L. (2017). Magnitude, precision, and realism of depth perception in stereoscopic vision. *Cognitive Research: Principles and Implications*, *2*(1), 25. https://doi.org/10.1186/s41235-017-0062-7

Hibbard, P. B., Goutcher, R., Hornsey, R. L., Hunter, D. W., & Scarfe, P. (2023, February). Luminance contrast provides metric depth information. *Royal Society Open Science*, *10*(2). https://doi.org/10.1098/rsos.220567

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 1–1. https://doi.org/10.1167/4.12.1

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), 33–33. https://doi.org/10.1167/8.3.33

Howard, I. P., & Rogers, B. J. (2012). *Perceiving in Depth, Volume 2: Stereoscopic Vision*. Oxford University Press. https://books.google.co.uk/books?id=AG2JAgAAQBAJ

Howarth P. A. (2011). Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: a review. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, *31*(2), 111–122. https://doi.org/10.1111/j.1475-1313.2011.00822.x

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629. https://doi.org/10.1016/S0042-6989(99)00088-7

Jorke, H., Simon, A., & Fritz, M. (2009). Advanced stereo projection using interference filters. *Journal of the Society for Information Display*, *17*(5), 407–410. https://doi.org/10.1889/JSID17.5.407

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. https://doi.org/10.1016/S0042-6989(03)00458-9

Kooi, F. L., & Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays*, *25*(2), 99–108. https://doi.org/10.1016/j.displa.2004.07.004

Kotulak, J. C., & Schor, C. M. (1986). A computational model of the error detector of human visual accommodation. *Biological cybernetics*, *54*(3), 189–194. https://doi.org/10.1007/BF00356857

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Krishnan, V. V., Shirachi, D., & Stark, L. (1977). Dynamic measures of vergence accommodation. *American journal of optometry and physiological optics*, *54*(7), 470–473. https://doi.org/10.1097/00006324-197707000-00007

Kruger, P. B., Mathews, S., Aggarwala, K. R., Yager, D., & Kruger, E. S. (1995). Accommodation responds to changing contrast of long, middle and short spectral-waveband components of the retinal image. *Vision Research*, *35*(17), 2415–2429. https://doi.org/10.1016/0042-6989(94)00316-5

Künnapas, T. (1968). Distance perception as a function of available visual cues. Journal of Experimental Psychology, 77, 523-529.

Lambooij, M., Fortuin, M., Heynderickx, I., IJsselsteijn, W., Fortuin, M., Heynderickx, I., & IJsselsteijn, W. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology*, *53*, 1–14. https://doi.org/10.2352/J.ImagingSci.Technol.2009.53.3.030201

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, *18*(9), 2307–2320. https://doi.org/10.1364/JOSAA.18.002307

Laver, K., George, S., Thomas, S., Deutsch, J., & Crotty, M. (2015). Virtual reality for stroke rehabilitation: An abridged version of a Cochrane review. *European Journal of Physical and Rehabilitation Medicine*, *51*(4), 497—506.

Lee, J. H., Stark, L. R., Cohen, S., & Kruger, P. B. (1999). Accommodation to static chromatic simulations of blurred retinal images. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, *19*(3), 223–235. https://doi.org/10.1046/j.1475-1313.1999.00440.x

Leigh, R. J., & Zee, D. S. (2015). The Neurology of Eye Movements. Oxford University Press.

Liu, S., Cheng, D., & Hua, H. (2008). An Optical See-through Head Mounted Display with Addressable Focal Planes. *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, 33–42. https://doi.org/10.1109/ISMAR.2008.4637321

Love, G. D., Hoffman, D. M., Hands, P. J. W., Gao, J., Kirby, A. K., & Banks, M. S. (2009). High-speed switchable lens enables the development of a volumetric stereoscopic display. *Opt. Express*, *17*(18), 15716–15725. https://doi.org/10.1364/OE.17.015716

MacKenzie, K. J., Hoffman, D. M., & Watt, S. J. (2010). Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision*, *10*(8), 22–22. https://doi.org/10.1167/10.8.22

Martens, T. G., & Ogle, K. N. (1959). Observations on accommodative convergence; especially its nonlinear relationships. *American journal of ophthalmology*, *47*(1 Pt 2), 455–463.

Mather G. (1997). The use of image blur as a depth cue. *Perception*, *26*(9), 1147–1158. https://doi.org/10.1068/p261147

Mather, G., & Smith, D. R. (2000). Depth cue integration: stereopsis and image blur. *Vision research*, *40*(25), 3501–3506. https://doi.org/10.1016/s0042-6989(00)00178-4

Mather, G., & Smith, D. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception*, *31*(10), 1211–1219. https://doi.org/10.1068/p3254

Nefs, H. T. (2012). Depth of Field Affects Perceived Depth-width Ratios in Photographs of Natural Scenes. *Seeing and Perceiving*, *25*(6), 577–595. https://doi.org/10.1163/18784763-00002400

Nguyen, V. A., Howard, I. P., & Allison, R. S. (2005). Detection of the depth order of defocused images. *Vision Research*, *45*(8), 1003–1011. https://doi.org/10.1016/j.visres.2004.10.015

O'Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research*, *34*(12), 1595–1604. https://doi.org/10.1016/0042-6989(94)90116-3

Palmer, Stephen (1999). Vision Science: Photons to Phenomenology. MIT Press.

Pastoor, S., & Wöpking, M. (1997). 3-D displays: A review of current technologies. *Displays*, *17*(2), 100–110. https://doi.org/10.1016/S0141-9382(96)01040-2

Perez-Marcos, D., Chevalley, O., Schmidlin, T., Garipelli, G., Serino, A., Vuadens, P., Tadi, T., Blanke, O., & Millán, J. d. R. (2017). Increasing upper limb training intensity in chronic stroke using embodied virtual reality: A pilot study. *Journal of NeuroEngineering and Rehabilitation*, *14*(1), 119. https://doi.org/10.1186/s12984-017-0328-9

Percival A. S. (1892). The relation of convergence to accommodation and its practical bearing. Ophthalmological Review, 11, 313–328.

Pizlo, Z., & Scheessele, M. (1998). Perception of 3D scenes from pictures. Proc SPIE. https://doi.org/10.1117/12.320131

Remington, L. A. (2012). Chapter 1—Visual System. In L. A. Remington (Ed.), *Clinical Anatomy and Physiology of the Visual System (Third Edition)* (pp. 1–9). Butterworth-Heinemann. https://doi.org/10.1016/B978-1-4377-1926-0.10001-3

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, *22*(2), 261–270. https://doi.org/10.1016/0042-6989(82)90126-2

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, *361*(6409), 253–255. https://doi.org/10.1038/361253a0

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, *24*(2), 155–179. https://doi.org/10.1068/p240155

Rogers, B. (2019). Toward a new theory of stereopsis: A critique of Vishwanath (2014). *Psychological Review, 126*(1), 162–169. https://doi.org/10.1037/rev0000131

Rolland, J. P., Krueger, M. W., & Goon, A. A. (1999). Dynamic focusing in head-mounted displays. *Electronic Imaging*.

Saladin, J. J., & Sheedy, J. E. (1978). Population Study of Fixation Disparity, Heterophoria, and Vergence. *Optometry and Vision Science*, *55*(11).

Semmlow, J., & Wetzel, P. (1979). Dynamic contributions of the components of binocular vergence. *Journal of the Optical Society of America*, *69*(5), 639–645. https://doi.org/10.1364/josa.69.000639

Schechner, Y. Y., & Kiryati, N. (2000). Depth from Defocus vs. Stereo: How Different Really Are They? *International Journal of Computer Vision*, *39*(2), 141–162. https://doi.org/10.1023/A:1008175127327

Schor, C. M., & Tsuetaki, T. K. (1987). Fatigue of accommodation and vergence modifies their mutual interactions. *Investigative ophthalmology & visual science*, *28*(8), 1250–1259.

Sexton, I., & Surman, P. (1999). Stereoscopic and autostereoscopic display systems. *IEEE Signal Processing Magazine*, *16*(3), 85–99. https://doi.org/10.1109/79.768575

Sheard, C. (1934). The prescription of prisms: As determined by analyses of data on relative amplitudes of convergence and accommodation. *Optometry and Vision Science*, *11*(10), 364.

Sheedy, J. E., Hayes, J. N., & Engle, J. (2003). Is all asthenopia the same?. *Optometry and vision science : official publication of the American Academy of Optometry*, *80*(11), 732–739. https://doi.org/10.1097/00006324-200311000-00008

Shevlin, F. (2005). *A fixed-viewpoint volumetric stereoscopic 3D display using adaptive optics*. *5664*, 22–27. https://doi.org/10.1117/12.585913

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, *11*(8), 11–11. https://doi.org/10.1167/11.8.11

Silva, J. N. A., Southworth, M., Raptis, C., & Silva, J. (2018). Emerging Applications of Virtual Reality in Cardiovascular Medicine. *JACC: Basic to Translational Science*, *3*(3), 420–430. https://doi.org/10.1016/j.jacbts.2017.11.009

Simpson, A. L., Sun, K., Pheiffer, T. S., Rucker, D. C., Sills, A. K., Thompson, R. C., & Miga, M. I. (2014). Evaluation of Conoscopic Holography for Estimating Tumor Resection Cavities in Model-Based Image-Guided Neurosurgery. *IEEE Transactions on Biomedical Engineering*, *61*(6), 1833–1843. https://doi.org/10.1109/TBME.2014.2308299

Suyama, S., Date, M., & Takada, H. (2000). Three-Dimensional Display System with Dual-Frequency Liquid-Crystal Varifocal Lens. *Japanese Journal of Applied Physics*, *39*(2R), 480. https://doi.org/10.1143/JJAP.39.480

Thatte, J. (2020). Cinematic Virtual Reality with Head-Motion Parallax [Ph.D., Stanford University]. In *ProQuest Dissertations and Theses* (2509609963). ProQuest Dissertations & Theses Global.

Trentacoste, M., Mantiuk, R., & Heidrich, W. (2011). Blur-Aware Image Downsampling. *Computer Graphics Forum*, *30*(2), 573–582. https://doi.org/10.1111/j.1467-8659.2011.01894.x

Ukai, K., & Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Health and Safety Aspects of Visual Displays*, *29*(2), 106–116. https://doi.org/10.1016/j.displa.2007.09.004

Urvoy, M., Barkowsky, M., & Le Callet, P. (2013). How visual fatigue and discomfort impact 3D-TV quality of experience: A comprehensive review of technological, psychophysical, and psychological factors. *Annals of Telecommunications - Annales Des Télécommunications*, *68*(11), 641–655. https://doi.org/10.1007/s12243-013-0394-3 van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, *122*(4), 367–377. https://doi.org/10.1007/s002210050525

Vienne, C., Plantier, J., Neveu, P., & Priot, A.-E. (2018). (Disparity-Driven) Accommodation Response Contributes to Perceived Depth. *Frontiers in Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnins.2018.00973

Vinnikov, M., Allison, R., & Fernandes, S. (2016). Impact of Depth of Field Simulation on Visual Fatigue: Who are Impacted? And How? *International Journal of Human-Computer Studies*, *91*. https://doi.org/10.1016/j.ijhcs.2016.03.001

Vishwanath, D., & Blaser, E. (2010). Retinal blur and the perception of egocentric distance. *Journal of Vision*, *10*(10), 26–26. https://doi.org/10.1167/10.10.26

Vishwanath, D. (2020). Advancing a new theory of stereopsis: Reply to Rogers (2019). *Psychological Review, 127*(1), 146–152. https://doi.org/10.1037/rev0000168

Wann, J. P., & Mon-Williams, M. (1997). Health Issues with Virtual Reality Displays: What We Do Know and What We Don't. *SIGGRAPH Comput. Graph.*, *31*(2), 53–57. https://doi.org/10.1145/271283.271307

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005) (**a**). Focus cues affect perceived depth. *Journal of Vision*, *5*(10), 7–7. https://doi.org/10.1167/5.10.7

Watt, S., Akeley, K., Girshick, A., & Banks, M. (2005) (**b**). Achieving near-correct focus cues in a 3-D display using multiple image planes. *Proceedings of SPIE - The International Society for Optical Engineering*, *5666*. https://doi.org/10.1117/12.610851

Wheatstone, C. (1838). 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. JSTOR.

Zhang, T., O'hare, L., Hibbard, P. B., Nefs, H. T., & Heynderickx, I. (2014). Depth of Field Affects Perceived Depth in Stereographs. *ACM Trans. Appl. Percept.*, *11*(4). https://doi.org/10.1145/2667227

Zhong, F., Jindal, A., Yöntem, A. Ö., Hanji, P., Watt, S. J., & Mantiuk, R. K. (2021). Reproducing Reality with a High-Dynamic-Range Multi-Focal Stereo Display. *ACM Trans. Graph.*, *40*(6). https://doi.org/10.1145/3478513.3480513