# Perceptual Requirements for Eye-Tracked Distortion Correction in VR

Phillip Guan
Reality Labs Research, Meta
United States of America
philguan@fb.com

Olivier Mercier
Reality Labs Research, Meta
United States of America
omercier@fb.com

Michael Shvartsman
Reality Labs Research, Meta
United States of America
michael.shvartsman@fb.com

Douglas Lanman
Reality Labs Research, Meta
United States of America
douglas.lanman@fb.com

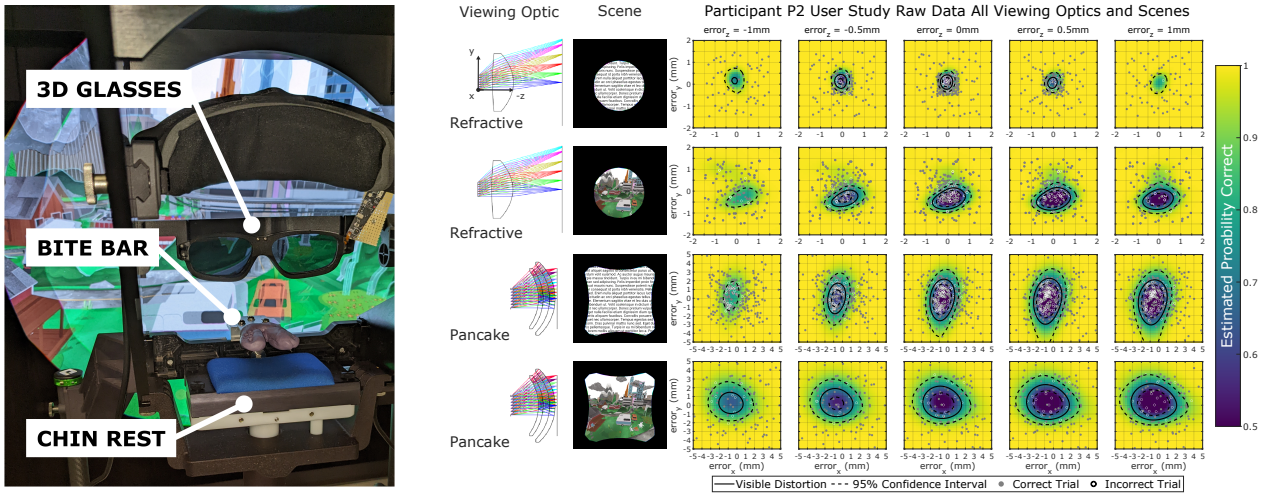


**Figure 1: (Left) A virtual reality (VR) display simulator that emulates the optical distortions produced by any type of viewing optic. (Right) Representative study results. Data is collected using the *AEPsych* framework [Owen et al. 2021], which accelerates high-dimensional psychophysical studies. As a result, we report and analyze the first user studies of VR distortion correction that examine the joint requirements for eye tracking bias in three dimensions and latency.**

## ABSTRACT

We present a virtual reality display system simulator that accurately reproduces gaze-contingent distortions created by any viewing optic. The simulator hardware supports rapid prototyping by presenting stereoscopic distortions on a high-speed television paired with shutter glasses, eliminating the need to fabricate physical optics. We further introduce light field portals as an efficient and general-purpose representation for VR optics, enabling real-time emulation using our simulator. This platform is used to conduct the first user study of perceptual requirements for eye-tracked optical distortion correction. Because our hardware platform facilitates consistent head and eye movements, it enables direct comparison of these requirements across observers, optical designs, and scene content. We conclude by introducing a simple binocular distortion metric, built using light field portals, which agrees with key trends identified in the user study and lays a foundation for the design of perceptually-based distortion metrics and correction schemes.

## CCS CONCEPTS

• **Computing methodologies** → **Virtual reality**.

## KEYWORDS

distortion correction, pupil swim, eye tracking

**ACM Reference Format:**
Phillip Guan, Olivier Mercier, Michael Shvartsman, and Douglas Lanman. 2022. Perceptual Requirements for Eye-Tracked Distortion Correction in VR. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings (SIGGRAPH '22 Conference Proceedings), August 7–11, 2022, Vancouver, BC, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3528233.3530699


*SIGGRAPH '22 Conference Proceedings, August 7–11, 2022, Vancouver, BC, Canada*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/3528233.3530699

# 1 INTRODUCTION

Virtual reality (VR) headsets offer a uniquely immersive visual experience, enabled by modern near-eye viewing optics with compact form factors and wide fields of view (FOV) [Koulieris et al. 2019]. However, these optics produce a number of visual artifacts that are not seen with direct-view displays, such as significant optical distortion. In practice, this is mitigated by warping the displayed image, resulting in the barrel-distorted imagery synonymous with VR [Robinett and Rolland 1993; Rolland and Hopkins 1993]. While such distortion correction generally straightens otherwise distorted lines, it only provides complete elimination of residual optical distortions when the entrance pupil of the eye is at the nominal, designed location. If the rendered distortion correction does not update with eye movement, the apparent optical distortion will continually change—a phenomenon known as pupil swim [Geng et al. 2018].

The user impact of pupil swim is fundamentally a perceptual question: what type and magnitude of residual optical distortions are imperceptible? We note that visual sensitivity to motion is higher than visual acuity [Mckee and Nakayama 1984]; thus, it is the *dynamic* nature of pupil swim that may make it particularly noticeable. In theory, eye-tracked dynamic distortion correction (DDC) could be applied, updating the rendered distortion correction to account for movement of the eyes, as reported by an eye tracker. However, DDC presents practical challenges, as neither the eye tracking system nor the optical distortion model will be fully accurate (e.g., due to manufacturing errors). As a result, residual, uncorrected distortions will exist in practice. While a geometric analysis can determine the residual distortion for a given level of eye tracking and optical modeling accuracy, such analysis alone is insufficient to determine the = of these distortions.

Evaluating the perception of optical distortions in VR headsets is difficult for many reasons. First, pupil swim is induced by user eye movements, which are usually accompanied by head and body movements. Because visual-vestibular cue integration influences how visual cues are interpreted [Butler et al. 2010; Cuturi and MacNeilage 2014; Fetsch et al. 2009; MacNeilage et al. 2012], it is important to design experiments with active observers to study pupil swim in an ecologically valid way (i.e., allowing for some degree of natural eye and head movements). Second, visual signals in the periphery play an important role in how distortions are perceived [Allison et al. 1999], so stimuli should be presented to match wide VR FOV conditions. Third, because pupil swim results from viewing displays through physical optics, it is challenging to control or even characterize the distortion seen through real optics when the user's entrance pupil positions are unknown. Notably, determining the entrance pupil positions is even more challenging when the observer moves their head and eyes.

For these reasons, we identify the need to create a testbed allowing controlled user studies of VR display systems with active observers. Previously, others have relied on the optics within existing headsets to produce distortions, and then varied the distortion correction parameters—using depth, shape, and distance judgments to illustrate the consequences of incomplete distortion correction [Hornsey et al. 2020; Knapp and Loomis 2003; Kuhl et al. 2008; Tong et al. 2019, 2020]. However, the residual distortion observed will depend on headset fit, which can be highly variable. Furthermore, study tasks ask subjects to provide a single, global judgment to characterize an artifact that has significant spatial and temporal variations. Correspondingly, performance on such tasks is not accurate, even when employing direct-view displays [Bradshaw et al. 1996; Cumming et al. 1991; Guan and Banks 2016; Hartle and Wilcox 2022; Johnston 1991; Watt et al. 2005]. In this work, we make the following contributions to address these concerns.

- We build a VR display simulator that accurately reproduces optical distortions, bypassing the need to physically construct lenses to evaluate their distortion properties.
- We introduce *light field portals* (LFPs) as an efficient representation for VR viewing optics, enabling real-time emulation and correction of gaze-contingent lens distortions.
- We conduct the first psychophysical study of eye tracking requirements for dynamic distortion correction (DDC). Specifically, we study user tolerances for eye tracking bias and latency, highlighting variations in detection thresholds across different participants, lens designs, and scene content.
- We analyze the study data and present a geometric model, enabled by LFPs, that agrees with observed trends.

# 2 RELATED WORK

*Characterizing VR Viewing Optics.* Misalignment between the eye's entrance pupil and the optical axis of a VR lens is akin to lens decentration in cameras [Brown 1966]. In previous work, Jones et al. [2015] characterize these distortions using computer vision methods. In closely related work, Geng et al. [2018] evaluate how pupil swim geometrically varies for three VR lenses: a smooth refractive, a Fresnel, and a polarization-based pancake. Similarly, Cakmakci et al. [2019] and Cholewiak et al. [2020] model gaze-contingent lens properties to define a perceptual eyebox. These works illustrate how to apply geometric analysis to understand the perceptual implications of various optical design choices. However, our work is the first to report detailed user studies of VR lens distortions, using the first *binocular* model to analyze our results.

*Distortion Correction Algorithms.* Static distortion correction for VR optics is a well established technique [Robinett and Rolland 1993; Rolland and Hopkins 1993]. Martschinke et al. [2019] implement dynamic distortion correction in an Oculus Rift DK2 with a Pupil Labs eye tracker. However, their approach is not able to represent a high-fidelity correction that varies with entrance pupil position, nor does their implementation account for changes in eye relief across different individuals or during eye rotations. Other works like [Hullin et al. 2012] and [Schrade et al. 2016] can account for varying pupil positions, but use only a few coefficients to represent lens distortions (which can have thousands of coefficents). Our LFP representation can use 100,000s of coefficients to represent the lens distortion to high accuracies while maintaining real-time frame rates and allows for a more general approach to aperture sampling.

*Perceptual Outcomes for Incomplete Distortion Correction.* Pupil swim is caused by a mismatch between the distortion correction applied and the distortion correction required. A small number of studies have explicitly manipulated distortion correction in VR and found conflicting results. Kuhl et al. [2008] manipulate the degree of pincushion distortion and find no effect on distance estimation

from three to six meters. Tong et al. [2019; 2020] manipulate distortion correction parameters and examine the effects of incomplete distortion on slant-angle estimation of planes, finding significant bias with inaccurate distortion correction. To our knowledge, our work is the first to explicitly evaluate the detectability of geometric distortions from pupil swim.

*Distortions from Rendering and Viewing Position Differences.* Like pupil swim, geometric distortions caused by offsets between the rendering cameras and the viewer's entrance pupils can be corrected with eye tracking. These perspective distortions are not the focus of our work, but they are closely related. The geometry and consequences of "ocular parallax" distortions are documented by Woods et al. [1993] and Held and Banks [2008]. Hwang and Peli [2019] examine the optical flow induced by rendering with incorrect camera parameters as the user moves. Perceptual studies by Pollack et al. [2012] measure the perceived depth of a rendered object when the rendering camera is offset from the correct center of projection. Wann et al. [1995] and Rolland et al. [2004] highlight potential distortions when a moving entrance pupil is not accounted for. Similarly, Krajancich et al. [2020] demonstrate improvements in depth constancy and depth alignment when the rendering cameras are placed at the entrance pupils, rather than the centers of rotation for the eyes (an offset of about 8mm). In our work we account for ocular parallax, correctly updating our rendering cameras to correspond with the entrance pupils of the user's eyes.

## 3 A VR DISPLAY SYSTEM SIMULATOR

This section and Section 2 of the supplementary materials describe the hardware and software elements used to build a VR display distortion simulator optimized for user studies. Our design is motivated by two key choices. First, we desire a means to prototype candidate VR viewing optics that does not require their physical fabrication, enabling user studies to rapidly inform optical design iterations. Second, we aim to facilitate accurate, repeatable vestibulo-ocular reflex (VOR) eye and head movements that serve as a controlled means to support active observers in studies.

### 3.1 Emulating VR Headsets with 3D TVs

Fabricating lenses can take weeks to months from the initial optical design. As an alternative, we *emulate* lenses using the simulator in Figure 2. Custom shutter glasses are synchronized with a large, high-speed OLED television (LG OLED88ZXPUA) to support stereoscopic viewing (resulting in a 60Hz refresh rate for each eye). The television is positioned 60cm away from the simulator's center of rotation, achieving a 125°×94° FOV and an angular resolution of 2.9 arcminutes[1,2] for the central pixel when the head is not rotated.

### 3.2 Why Study VOR Eye Movements?

As established in Sections 1 and 2, a complete investigation of pupil swim should incorporate active users controlling their own head and eye movements. However, the parameter space for quantifying optical distortions is too large to study without enforcing consistent movement across trials. We focus on VOR eye movements because they represent a worst case scenario for pupil swim; image features are stabilized on the retina while subjected to changing distortion, and sensitivity to visual-vestibular conflict during VOR is heightened relative to smooth pursuits [Garzorz and MacNeilage 2017]. Our testbed allows subjects to control their own VOR eye movements while also ensuring that the head and eye movements are consistent across trials. This is achieved by attaching a chin rest and bite bar to a rotary encoder so that the head always moves along a consistent arc.

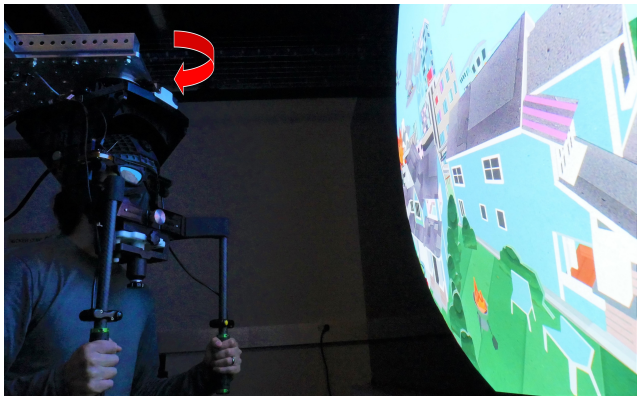

**Figure 2: Our VR display system simulator facilitates repeatable, accurately measured head and eye movements. Active shutter glasses are synchronized to a 120hz OLED display, enabling stereoscopic image presentation.**

### 3.3 Achieving Accurate Head and Eye Tracking

A key concern when allowing active head and eye movements is tracking accuracy. Latency, bias, and noise will result in dynamic errors during simulations and their magnitude may exceed that of the pupil swim distortions under consideration. The encoders in our testbed addresses this concern by precisely estimating the head rotation and translation from the associated mechanical encoder values (both reported at 1.0kHz with a latency <200$\mu$s). The entrance pupil positions are estimated by presenting a known fixation target and applying the eye model of Krajancich et al. [2020],which uses a 7.8mm offset from the eye's center of rotation to the entrance pupil. We further measure and compensate for the user's interpupillary distance using a pupilometer.

### 3.4 Light Field Portals

Standard rendering engines can reproduce perspective changes for tracked head and eye movements. However, to emulate optical distortions, we require a real-time method for simulating light transport through arbitrary viewing optics. Ray tracing for such complex optical systems is often too expensive to support interactive frame rates. Instead, we introduce *light field portals (LFPs)* as an efficient representation meeting our real-time rendering requirements.

As shown in Figure 3, we model any set of optical elements as a function that maps rays from an input plane to an output plane. Provided with this function, ray tracing can be used to compute the mapping between the visual angles from the entrance pupil to

[1]The instantaneous FOV and resolution vary with user head movement
[2]The native panel resolution is 7680x4320, so the true resolution is 1.45 arcminutes, but the display is limited to 3840x2160 when driven at 120Hz.

pixels on the display. For a given viewing direction, a ray starts at the entrance pupil and is traced to the input plane of the LFP. The ray is transformed by the LFP into the corresponding ray on the other side of the optical stack. The output ray is then traced in air until it intersects the display. To generate an entire image, the horizontal and vertical viewing angle space is divided into a mesh (128×128 triangles). Each mesh vertex establishes the eye-to-display mapping through the LFP, and each triangle is filled using interpolation from its three vertices. This mesh-based approach is easily implemented on standard GPU hardware.

Generating an LFP for a viewing optic requires a large number of ray traces (typically 100 million) to adequately sample the eyebox. These rays may be generated using most lens design applications (e.g., Zemax and Code V). The ray intersections with the input and output planes are recorded and the LFP representation is fitted (in the least-squares sense). In our implementation, the LFP is represented by 4D cubic Hermite polynomials defined on a low-resolution grid (e.g., 10×10×10×10 nodes), where each node requires 16 coefficients to account for all first-order derivatives. In practice, we find that LFP fitting errors are less than 1.4arcmin for each of the headset architectures we consider in this paper.

Using a rectilinear grid to represent generally round VR optics can lead to invalid regions of the data grid in the LFP, where input rays should be discarded instead of traced forward to the display. A level set function is computed on the same data grid to represent the valid region, essentially defining the view aperture of the optical stack in position and angle. The number of coefficients required to represent the LFP is

$$\begin{aligned}&(10^4 \text{ nodes}) \times (16 \text{ polynomial coefficients}) \times (4 \text{ outputs})\\ &+ (10^4 \text{ scalar coefficient for the valid level set})\\ &= 810,000 \text{ coefficients},\end{aligned}$$

which amounts to about 3MB of data. One LFP is used per color channel to account for chromatic aberration. Headsets are binocular, so approximately 18MB (3×3×2) are required to represent a complete headset with LFPs. By default, LFPs can represent architectures where the display moves relative to the optical stack (i.e., a varifocal architecture), but LFPs can be simplified for systems where the display is static. In this case, the output plane of the LFP is placed directly at the display, eliminating the need to trace the output LFP ray to the display and the $(\theta'_x, \theta'_y)$ output of the LFP can be ignored, reducing the number of stored coefficients by almost half. A parallel implementation of the LSQR method [Paige and Saunders 1982] on the GPU is used to fit the ray data to the LFP. Representations other than cubic Hermite polynomials could also be used, but the basis functions should have continuous first derivatives to avoid major artifacts in the distortion computations.

For a given lens design, the LFP representation is not necessarily the best choice in terms of memory usage and efficiency, and more specialized representations are often available. However, our LFP representation is versatile and well suited for our purposes since it can represent any reasonable lens system. This enables the rapid evaluation of new lens designs, regardless of their complexity, without changing our software pipeline.

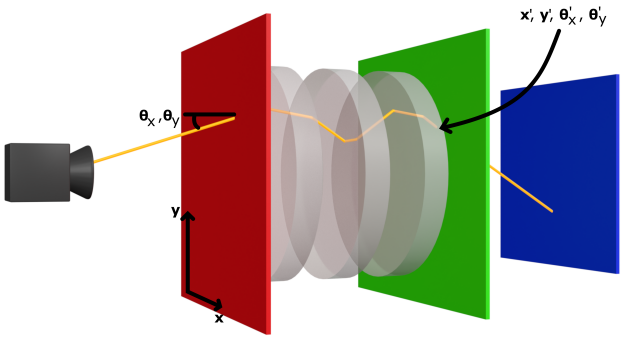


**Figure 3: A *light field portal (LFP)* represents any optical system as a 4D function. Instead of tracing rays through multiple lenses (gray), the LFP function maps a 4D ray from an input plane (red) to an output plane (green). The rays are propagated to a display (blue) to compute the mapping between view angles and display pixels for any pupil position.**

## 4 DISTORTION CORRECTION USER STUDY

This section describes a psychophysical study measuring the eye tracking requirements necessary to eliminate perceived pupil swim in a VR environment. Raw data and additional implementation details can be found in Section 1 of the supplementary materials.

### 4.1 Stimulus and Task

The stimulus used to present distortion is more complex than traditional low-level psychophysical stimuli (e.g., spatial frequency gratings), so higher variability across subjects is expected. To address this, we employed a more time-intensive, two-interval forced choice procedure, instead of a faster but more subjective yes/no judgment. In each interval participants were asked to fixate on a target in a text or 3D scene (Figure 1) during a horizontal VOR eye movement. Users controlled the speed, duration, and magnitude of their head rotation. They were not instructed to make their VOR movements with any particular magnitude or frequency. The scene in each interval was identical and both were rendered with DDC. However, one interval applied distortion correction with simulated eye tracking error and the other applied the correction without error; i.e., one interval contained distortion and the other was distortion-free. The two intervals were randomly ordered and subjects were asked to identify which interval contained distortion. When eye tracking errors led to distortions below a detectable threshold, subjects would identify the distorted interval at chance (50%) and at 100% when eye tracking error led to easily detectable distortion. Acceptable eye tracking performance was defined by the values of eye tracking error when distortion could be accurately identified 75% of the time.

### 4.2 A Nonparametric Model for Psychophysics

Traditional psychometric-function-based techniques, including adaptive methods such a QUEST [Watson and Pelli 1983] and QUEST+ [Watson 2017], can easily require 10 or more hours of data collection when applied to problems beyond three or more dimensions. This is because the total trials required to explore a problem space

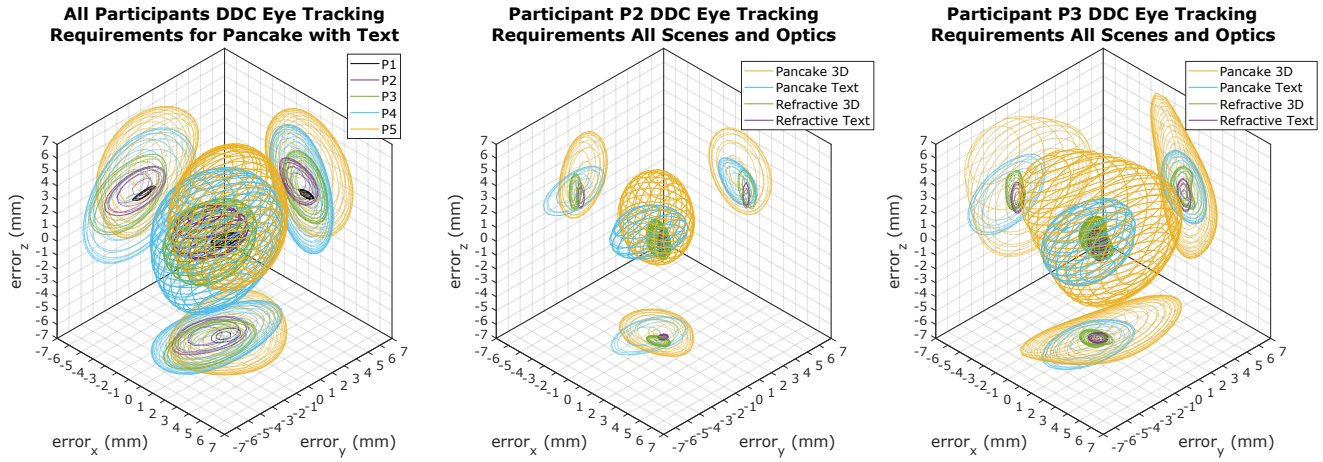


**Figure 4: (Left) Eye tracking requirements to eliminate perceptible distortions with dynamic distortion correction while viewing text with the pancake optic for five participants. Each surface describes the acceptable eye tracking bias for different observers in the x-, y-, and z-dimensions of the eyebox. (Center) Participant P2 DDC eye tracking requirements for text and 3D content viewed with pancake and refractive lenses. (Right) The same data for Participant P3.**

scales exponentially with the number of dimensions. A recent line of research on nonparametric models for the psychometric function [Gardner et al. 2015; Owen et al. 2021; Schlittenlacher et al. 2020, 2018; Song et al. 2017, 2018] provides a more efficient alternative. We utilized this approach by employing a Gaussian process model for the psychometric field from the *AEPsych* adaptive toolbox [Owen et al. 2021] and combined it with a space-filling Sobol quasi-random sequence [Sobol 1967] to guide data collection in the three dimensions of eye tracking bias for the studies in Sections 4.3 and 4.4. The combination of the space-filling design and flexible psychometric model enabled data capture for the entire 3D volume in as few as 299 trials compared to >2,500 trials using traditional methods. In Section 4.5 we used an adaptive sampling method developed by Letham et al. [2022] to map out continuous predictions over the full psychometric field for latency as a fourth dimension with just an additional 442 trials, a significant reduction from the estimated >10,000 trials required using traditional methods. Additional discussion on trial estimates and the use of *AEPsych* can be found in Section 1 of the supplementary materials.

## 4.3 Pancake Optic with Text

In the first experiment, six participants (two authors and four naive subjects, ages 23-33) performed the task in Section 4.1, for a polarization-based folded pancake optical design (Figure 1), while viewing text displayed on a plane coincident with the television. Specifically, we assessed the pancake lens published by Geng et al. [2018] (see their Figure 2). Each participant collected 299-909 trials, with the extracted detection thresholds shown on the left of Figure 4 (see the supplementary materials for additional results). Participants 3 and 6 (P3 and P6) usually wear eyeglasses, but did not wear them to avoid pupil swim from their own corrective lenses. Without their glasses, P3 had 20/30 vision, P6 had 20/40 vision; all other subjects had 20/20 vision. Likely due to reduced visual acuity, P6's thresholds are significantly larger than others' (thus, their data is excluded in Figure 4, but is included in the supplementary materials). Significant variability exists between subjects: P1 exhibits the most sensitivity to eye tracking errors, and P5 and P6 show the least. Interestingly, all subjects are less sensitive to errors along the y-axis compared to the x-axis. Several observers are also more sensitive to eye relief (z-axis) errors towards the display compared to away. We explore some of these effects in more detail in Section 5.

## 4.4 Impacts of Optics and Scene Content

Next we examine the impact of different optical designs and scenes. The study from Section 4.3 was repeated using the smooth refractive lens published by Geng et al. [2018]. While pancake lenses apply a secondary reflective surface to minimize pupil swim, refractive lenses contain only two optical surfaces and, correspondingly, generally exhibit larger pupil swim. Thus, eye tracking errors are expected to more negatively impact DDC for refractive lenses, but to an unknown perceptual extent. In this study, we further considered a 3D scene in addition to the text scene (see Figure 1). Participants P2 and P3 were subjects in this study and used a bite bar while viewing the 3D scene to further stabilize their heads during VOR movements, minimizing ocular parallax errors (Section 2). The results are shown in the two plots on the right of Figure 4. The varying responses to different optical designs and scene content are readily apparent for both participants. Eye tracking requirements are significantly relaxed for the pancake design (yellow and blue contours) compared to the refractive lens (green and purple contours). For a given optical design, eye tracking requirements are less stringent for 3D content (yellow and green contours) relative to text (blue and purple contours) for both participants.

## 4.5 Eye Tracking Latency Requirements

In addition to inaccuracy in the x-, y-, and z-dimensions, eye tracking error can also be introduced through latency. We simulated eye tracking latency by storing previous head rotation values from

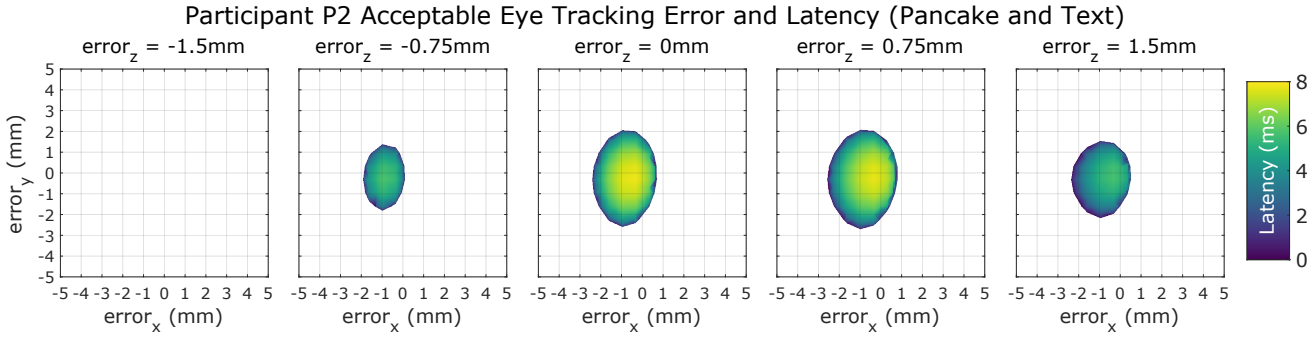


**Figure 5: Eye tracking requirements to eliminate perceptible distortions with dynamic distortion correction while viewing text with the pancake optic for P2. Each plot shows the acceptable latency for a given amount of eye tracking bias in the x-, y-, and z-dimensions of the eyebox. Each column represents a fixed amount of entrance pupil position bias in depth. Plots with the largest shaded regions represent conditions with the least stringent eye tracking requirements.**

the 1,000Hz rotary encoder in a rolling buffer and computing eye positions using a delayed head rotation during the VOR movement. Introducing latency as a fourth dimension to our study would have added additional inefficiency when using a space-filling Sobol generator, as a larger volume of the problem space is either well-above or well-below threshold. We instead used a model-based adaptive sampling method called *EAVC* [Letham et al. 2022] to increase sampling efficiency. In Figure 5, 442 points were collected using this adaptive method and combined with P2's 909 trials from Figure 4. Each column represents different amounts of eye tracking bias along the z-axis for the pancake lens while viewing text. The shaded region in each individual plot represents the combination of eye tracking latency and bias along the x- and y-axes and a fixed value of eye tracking bias along the z-axis, where DDC with eye tracking error is perceptually indistinguishable from a distortion-free stimulus. White regions in each plot represent combinations of eye tracking bias in x, y, and z, where DDC with those values results in distortions that are always visible. As expected, larger amounts of latency are tolerated when the eye tracking errors are smallest. In the best case scenario, when eye tracking errors are zero in all dimensions, latency as small as 8ms is detectable. Note that this value represents additional latency added to our system's inherent latency.

## 5 MODELING USER STUDY RESULTS

This section analyzes key trends from the user study using a new geometric model and associated distortion metric (Figure 6). Importantly, this metric considers the differences in distortion introduced to the left and right eyes as well as the underlying geometry of the content being rendered. The metric is built with our LFP framework and compares the intended 3D geometry of a scene to the geometry observed when emulating DDC with eye tracking errors. The distortion metric is a simple sum of the differences in binocular disparity in these two conditions over the entire set of images observed during a head and eye movement specified in the rendering engine (in this case a ±25° horizontal VOR with a horizontal offset of 8.5cm from the center of rotation). To compute the error, the scene is first rendered from the viewpoint centered between the eyes. Each pixel from the cyclopean view is projected into its corresponding 3D scene point and from there, each point is projected into the user's left and right views. By comparing the binocular disparity of each 3D point, with and without eye tracking errors, we produce an error map across the field of view. The final metric is the averaged binocular disparity difference across the field of view, expressed in arcminutes. Many additions to emulate the human visual system could be added to the metric, but even this basic geometric model predicts some key trends observed in Section 4.

*Asymmetries in Eye Relief.* The acceptable range of eye tracking errors in x and y is considerably smaller for several observers when eye relief errors place the eye towards the optics. For example, in the second and third rows of Figure 1, the contours representing acceptable error are smaller for negative values of eye relief error. The same trend can be seen for additional observers in Figure 1 of the supplementary materials. The bottom left of Figure 6 shows corresponding simulations using our binocular metric with eye relief error. For a fixed eye relief error, the metric is integrated over a ±25° VOR movement with x and y eye tracking errors within ±2mm. The metric minimum is shifted towards positive values of eye relief error and increases faster in the negative direction. Thus, both the user study and metric indicate that eye relief errors have an asymmetric impact on DDC for the pancake optic.

*Anisotropy in X and Y.* Figure 1 also indicates that acceptable eye tracking error contours are more elongated in y than in x for participant P2, a trend which is mirrored by several other participants in Figure 1 of the supplementary materials. Figure 6 shows our binocular metric, but instead aggregated over eye tracking errors fixed along the x or y axes. For a given value of horizontal eye tracking error we set the y-axis and eye relief errors to 0 and integrate the metric over a ±25° VOR movement. We use the same process for the y-axis error plot, but with the roles of x and y reversed. Like the eye relief analysis, the prediction of the binocular metric agrees with the trends identified in the user study.

*Differences in Optical Design.* For both P2 and P3, the pancake optic can tolerate more eye tracking error for DDC compared to the smooth refractive lens (center and right panels of Figure 4). Using our binocular metric, we sum the overall distortion for the two lens

designs as x, y, and z eye tracking errors vary over ±2mm for a VOR motion of ±25°. The metric is 8.48 arcmin for the smooth refractive lens and 5.74 arcmin for the pancake lens. This indicates that the same eye tracking errors for DDC lead to more error when applied to the refractive optic, which again agrees with the user study data. This result is expected as the pancake lens has properties which allow its design to reduce pupil swim [Geng et al. 2018].

*Effects of Scene Content.* Figure 1 and Figure 4 show that scene content has a significant impact on measured user tolerances to DDC eye tracking errors. However, when comparing our binocular metric for the smooth refractive lens between the two scenes in the user study, we compute a metric of 8.48 arcmin for the 3D scene and 8.10 arcmin for the text scene, which is negligible. This result indicates a likely limitation of our metric. Even though the scene content is part of the metric formulation, the model fails to reproduce this trend we find in the user study data. A necessary direction for future work is to develop models that better assess the effect of scene content on pupil swim perception, for example, to better account for the spatial and temporal properties of the introduced disparity errors.

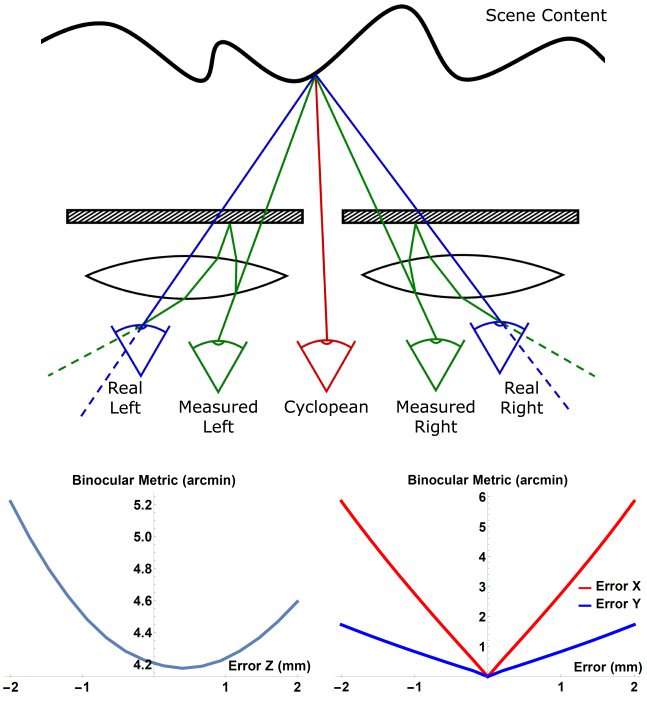


**Figure 6: (Top) Binocular distortion metric. The scene is rendered from a cyclopean camera (red). Pixels are reprojected to 3D and then projected into the user's left and right eyes. Using the LFPs for each lens, we compute the disparity that the viewer will perceive for each point when viewing content rendered using DDC with and without tracking errors (green and blue lines respectively). We define our distortion metric as the average binocular disparity difference between the two conditions. (Bottom Left) Metric for eye tracking errors in z. (Bottom Right) Metric for eye tracking errors in x and y.**

## 6 CONCLUSION AND FUTURE WORK

We built a binocular VR display system simulator for rapid prototyping of viewing optics that also serves as a user study testbed; designed and ran the first user study protocol to directly measure the acceptable eye tracking errors for DDC and found differences across observers, optical designs, and scenes; and, finally, introduced a binocular metric that agrees with many trends observed in the user study. We conclude by reviewing current limitations of our methodology and potential avenues of future work.

*User Study Limitations.* A small number of observers participated in the user studies (from n=1 to n=6). While the observed trends are consistent across observers, they are specific to the optical architectures and scene content presented. We believe these results motivate follow-up work that is required to understand the perceptual mechanisms underpinning these differences. Crucially, once understood, these mechanisms may be applied to drive perceptually optimized approaches for the design of VR optics and associated distortion correction algorithms.

*Hardware Limitations.* The OLED display used in the testbed offers superior resolution, contrast, and brightness, compared to any commercial VR headset. However, employing a television results in certain practical limitations for our simulator, including: potential clipping of the rendered field of view as the user's head rotates away from display; a variable level of crosstalk observed with the shutter glasses (which depends on the rotation angle of the head relative to the display); temporal artifacts due to time-multiplexed presentation of stereoscopic content at a limited refresh rate; and changes in focal distance and resolution as the head moves.

*Hardware Upgrades.* While our simulator is designed for rapid prototyping, some of the resulting limitations could be addressed by mounting VR headsets directly to our chin rest and bite bar mechanism. In this manner, the system could still accurately estimate the user's head and eye positions, while allowing studies to be conducted on the specific display system under consideration. However, such an approach would eliminate the ability to present "distortion-free" imagery unless the distortions of the optics were compensated for in real-time (e.g., using LFPs).

An eye tracker could be mounted to the shutter glasses in our simulator allowing for direct comparison of the eye tracking signal with the reference eye position produced by our testbed (i.e., as estimated by the model applied in our user study when presenting a known fixation target and measuring the head rotation). In this way, our simulator may provide a platform to characterize performance for existing and emerging eye trackers. The eye tracker could also be used to conduct user studies with free gaze, further expanding the prototyping and development framework of our simulator for VR display systems.

## ACKNOWLEDGMENTS

We thank Joel Hegland, Ryan Ebert, Dave Lindberg, Sam Pepperwood, Robert Birch, Chris Neugebauer, and Bruce Cleary for engineering support of the VR simulator, Nathan Matsuda for work on the custom shutter glasses, Marina Zannoli for feedback on experimental and testbed design, Ben Letham for assistance with the AEPsych platform, and Julia Majors for copyediting.

# REFERENCES


Robert S Allison, Ian P Howard, and James E Zacher. 1999. Effect of Field Size, Head Motion, and Rotational Velocity on Roll Vection and Illusory Self-Tilt in a Tumbling Room. *Perception* 28, 3 (1999), 299–306. https://doi.org/10.1068/p2891

Mark F. Bradshaw, Andrew Glennerster, and Brian J. Rogers. 1996. The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research* 36, 9 (1996), 1255–1264. https://doi.org/10.1016/0042-6989(95)00190-5

Duane C. Brown. 1966. Decentering distortion of lenses. In *Photogrammetric Engineering*, Vol. 32. 444–462.

John S. Butler, Stuart T. Smith, Jennifer L. Campos, and Heinrich H. Bülthoff. 2010. Bayesian integration of visual and vestibular signals for heading. *Journal of Vision* 10, 11 (09 2010), 23–23. https://doi.org/10.1167/10.11.23

Ozan Cakmakci, David M. Hoffman, and Nikhil Balram. 2019. 3D Eyebox in Augmented and Virtual Reality Optics. *SID Symposium Digest of Technical Papers* 50, 1 (2019), 438–441. https://doi.org/10.1002/sdtp.12950

Steven A. Cholewiak, Zeynep Başgöze, Ozan Cakmakci, David M. Hoffman, and Emily A. Cooper. 2020. A perceptual eyebox for near-eye displays. *Opt. Express* 28, 25 (Dec 2020), 38008–38028. https://doi.org/10.1364/OE.408404

Bruce G. Cumming, Elizabeth B. Johnston, and Andrew J. Parker. 1991. Vertical disparities and perception of three-dimensional shape. *Nature* 349, 6308 (January 1991), 411—413. https://doi.org/10.1038/349411a0

Luigi F. Cuturi and Paul R. MacNeilage. 2014. Optic Flow Induces Nonvisual Self-Motion Aftereffects. *Current Biology* 24, 23 (2014), 2817–2821. https://doi.org/10.1016/j.cub.2014.10.015

Christopher R. Fetsch, Amanda H. Turner, Gregory C. DeAngelis, and Dora E. Angelaki. 2009. Dynamic Reweighting of Visual and Vestibular Cues during Self-Motion Perception. *Journal of Neuroscience* 29, 49 (2009), 15601–15612. https://doi.org/10.1523/JNEUROSCI.2574-09.2009

Jacob R. Gardner, Paul Upchurch, Matt J. Kusner, Yixuan Li, Kilian Q. Weinberger, Kavita Bala, and John E. Hopcroft. 2015. Deep Manifold Traversal: Changing Labels with Convolutional Features. (2015). arXiv:1511.06421 http://arxiv.org/abs/1511.06421

Isabelle T. Garzorz and Paul R. MacNeilage. 2017. Visual-Vestibular Conflict Detection Depends on Fixation. *Current Biology* 27, 18 (2017), 2856–2861.e4. https://doi.org/10.1016/j.cub.2017.08.011

Ying Geng, Jacques Gollier, Brian Wheelwright, Fenglin Peng, Yusufu Sulai, Brant Lewis, Ning Chan, Wai Sze Tiffany Lam, Alexander Fix, Douglas Lanman, Yijing Fu, Alexander Sohn, Brett Bryars, Nelson Cardenas, Youngshik Yoon, and Scott McEldowney. 2018. Viewing optics for immersive near-eye displays: pupil swim/size and weight/stray light. In *Digital Optics for Immersive Displays*, Vol. 10676. International Society for Optics and Photonics, SPIE, 19–35. https://doi.org/10.1117/12.2307671

Phillip Guan and Martin S. Banks. 2016. Stereoscopic depth constancy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371, 1697 (2016), 20150253–20150253. https://doi.org/10.1098/rstb.2015.0253

Brittney Hartle and Laurie M. Wilcox. 2022. Stereoscopic depth constancy for physical objects and their virtual counterparts. *Journal of Vision* 22, 4 (03 2022), 9–9. https://doi.org/10.1167/jov.22.4.9

Robert T. Held and Martin S. Banks. 2008. Misperceptions in Stereoscopic Displays: A Vision Science Perspective. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*. Association for Computing Machinery, New York, NY, USA, 23–32. https://doi.org/10.1145/1394281.1394285

Rebecca Hornsey, Paul Hibbard, and Peter Scarfe. 2020. Size and shape constancy in consumer virtual reality. *Behavior Research Methods* 52 (05 2020). https://doi.org/10.3758/s13428-019-01336-9

Matthias B Hullin, Johannes Hanika, and Wolfgang Heidrich. 2012. Polynomial Optics: A construction kit for efficient ray-tracing of lens systems. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 1375–1383.

Alex D. Hwang and Eli Peli. 2019. Stereoscopic Three-dimensional Optic Flow Distortions Caused by Mismatches Between Image Acquisition and Display Parameters. *Journal of Imaging Science and Technology* 63, 6 (2019), 60412–1–60412–7(7). https://doi.org/10/ghrnbp

Elizabeth B. Johnston. 1991. Systematic distortions of shape from stereopsis. *Vision Research* 31, 7 (1991), 1351–1360. https://doi.org/10.1016/0042-6989(91)90056-B

J. A. Jones, L. C. Dukes, D. M. Krum, M. T. Bolas, and L. F. Hodges. 2015. Correction of geometric distortions and the impact of eye position in virtual reality displays. In *2015 International Conference on Collaboration Technologies and Systems (CTS)*. 77–83. https://doi.org/10.1109/CTS.2015.7210403

Joshua Knapp and Jack Loomis. 2003. *Visual Perception of Egocentric Distance in Real and Virtual Environments*. Vol. 11. 21–46. https://doi.org/10.1201/9781410608888.pt1

G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. *Computer Graphics Forum* 38, 2 (2019), 493–519. https://doi.org/10.1111/cgf.13654

Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2020. Optimizing Depth Perception in Virtual and Augmented Reality through Gaze-Contingent Stereo Rendering. *ACM Trans. Graph.* 39, 6, Article 269 (Nov. 2020), 10 pages. https://doi.org/10.1145/3414685.3417820

Scott A. Kuhl, William B. Thompson, and Sarah H. Creem-Regehr. 2008. HMD Calibration and Its Effects on Distance Judgments. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization* (Los Angeles, California) *(APGV '08)*. Association for Computing Machinery, New York, NY, USA, 15–22. https://doi.org/10.1145/1394281.1394284

Benjamin Letham, Phillip Guan, Chase Tymms, Eytan Bakshy, and Michael Shvartsman. 2022. Look-Ahead Acquisition Functions for Bernoulli Level Set Estimation. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*. https://doi.org/10.48550/ARXIV.2203.09751

Paul R. MacNeilage, Zhou Zhang, Gregory C. DeAngelis, and Dora E. Angelaki. 2012. Vestibular Facilitation of Optic Flow Parsing. *PLOS ONE* 7, 7 (07 2012), 1–8. https://doi.org/10.1371/journal.pone.0040264

J. Martschinke, J. Martschinke, M. Stamminger, and F. Bauer. 2019. Gaze-Dependent Distortion Correction for Thick Lenses in HMDs. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1848–1851. https://doi.org/10.1109/VR.2019.8798107

Suzanne P. Mckee and Ken Nakayama. 1984. The detection of motion in the peripheral visual field. *Vision Research* 24, 1 (1984), 25–32. https://doi.org/10.1016/0042-6989(84)90140-8

Lucy Owen, Jonathan Browder, Benjamin Letham, Gideon Stocek, Chase Tymms, and Michael Shvartsman. 2021. Adaptive Nonparametric Psychophysics. arXiv:2104.09549 [stat.ME]

Christopher C Paige and Michael A Saunders. 1982. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)* 8, 1 (1982), 43–71.

Brice Pollock, Melissa Burton, Jonathan W. Kelly, Stephen Gilbert, and Eliot Winer. 2012. The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (2012), 581–588. https://doi.org/10.1109/TVCG.2012.58

Warren Robinett and Jannick P. Rolland. 1993. A Computational Model for the Stereoscopic Optics of a Head-Mounted Display. In *Virtual Reality Systems*. Academic Press, Boston, 51–75. https://doi.org/10.1016/B978-0-12-227748-1.50013-5

Jannick Rolland, Yonggang Ha, and Cali Fidopiastis. 2004. Albertian errors in head-mounted displays: I. Choice of eye-point location for a near- or far-field task visualization. *J. Opt. Soc. Am. A* 21, 6 (Jun 2004), 901–912. https://doi.org/10.1364/JOSAA.21.000901

Jannick P. Rolland and Terry Hopkins. 1993. *A Method of Computational Correction for Optical Distortion in Head-Mounted Displays*. Technical Report TR93-045. University of North Carolina at Chapel Hill. http://www.cs.unc.edu/techreports/93-045.pdf

Josef Schlittenlacher, Richard E. Turner, and Brian C.J. Moore. 2020. Application of Bayesian Active Learning to the Estimation of Auditory Filter Shapes Using the Notched-Noise Method. *Trends in Hearing* 24 (2020). https://doi.org/10.1177/2331216520952992

Josef Schlittenlacher, Richard E. Turner, and Brian C. J. Moore. 2018. Audiogram estimation using Bayesian active learning. *The Journal of the Acoustical Society of America* 144, 1 (2018), 421–430. https://doi.org/10.1121/1.5047436

Emanuel Schrade, Johannes Hanika, and Carsten Dachsbacher. 2016. Sparse high-degree polynomials for wide-angle lenses. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 89–97.

I.M Sobol. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *U. S. S. R. Comput. Math. and Math. Phys.* 7, 4 (jan 1967), 86–112. https://doi.org/10.1016/0041-5553(67)90144-9

Guoli Song, Shuhui Wang, Qingming Huang, and Qi Tian. 2017. Multimodal Similarity Gaussian Process Latent Variable Model. *IEEE Transactions on Image Processing* 26, 9 (2017), 4168–4181. https://doi.org/10.1109/TIP.2017.2713045

Xinyu D. Song, Kiron A. Sukesan, and Dennis L. Barbour. 2018. Bayesian active probabilistic classification for psychometric field estimation. *Attention, Perception, and Psychophysics* 80, 3 (2018), 798–812. https://doi.org/10.3758/s13414-017-1460-0

Jonathan Tong, Robert S. Allison, and Laurie M. Wilcox. 2019. The Impact of Radial Distortions in VR Headsets on Perceived Surface Slant. *Electronic Imaging, Human Vision and Electronic Imaging* 11 (2019), 60409–1–60409–11. https://doi.org/10.2352/J.ImagingSci.Technol.2019.63.6.060409

Jonathan Tong, Robert S. Allison, and Laurie M. Wilcox. 2020. Optical distortions in VR bias the perceived slant of moving surfaces. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 73–79. https://doi.org/doi:10/gktm46S

John P. Wann, Simon Rushton, and Mark Mon-Williams. 1995. Natural problems for stereoscopic depth perception in virtual environments. *Vision Research* 35, 19 (1995), 2731–2736. https://doi.org/10.1016/0042-6989(95)00018-U

Andrew B Watson. 2017. QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of vision (Charlottesville, Va.)* 17, 3 (2017), 10–10.

Andrew B. Watson and Denis G. Pelli. 1983. QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics* 33, 2 (1983), 113–120. https://doi.org/10.3758/BF03202828

Simon J. Watt, Kurt Akeley, Marc O. Ernst, and Martin S. Banks. 2005. Focus cues affect perceived depth. *Journal of Vision* 5, 10 (12 2005), 7–7. https://doi.org/10.1167/5.10.7

Andrew J. Woods, Tom Docherty, and Rolf Koch. 1993. Image distortions in stereoscopic video systems. In *Stereoscopic Displays and Applications IV*, Vol. 1915. International Society for Optics and Photonics, SPIE, 36–48. https://doi.org/10.1117/12.157041