To appear at ACM ETRA '14.

Comparing Estimated Gaze Depth in Virtual and Physical Environments

Andrew T. Duchowski[†]*, Donald H. House[†], Jordan Gestring[†], Robert Congdon[‡]
[†]School of Computing, [‡]Sonoco Institute of Packaging Design & Graphics
Clemson University

Lech Świrski, Neil A. Dodgson
Computer Laboratory
University of Cambridge

Krzysztof Krejtz[△κ], Izabela Krejtz*
[△]Department of Psychology, *Interdisciplinary Center for Applied Cognitive Studies
University of Social Sciences and Humanities,
[κ]Information Processing Institute

(a) Wheatstone-*style haploscope used for the virtual environment.*

(b) Dikablis/Vicon *system used for the physical environment.*

Figure 1: *Virtual and physical environments used in the study to compare vergence estimates.*

Abstract

We show that the error in 3D gaze depth (vergence) estimated from binocularly-tracked gaze disparity is related to the viewing distance of the screen calibration plane at which 2D gaze is recorded. In a stereoscopic (virtual) environment, this relationship is evident in gaze to target depth error: vergence error behind the screen is greater than in front of the screen and is lowest at the screen depth. In a physical environment, with no accommodation-vergence conflict, the magnitude of vergence error in front of the 2D calibration plane appears reversed, increasing with distance from the viewer.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality. J.4 [Computer Applications]: Social and Behavioral Sciences—Psychology.

Keywords: vergence movements, head-mounted systems, stereoscopic displays

*duchowski@clemson.edu

ETRA '14, March 26–28 2014, Safety Harbor, FL, USA

http://dx.doi.org/10.1145/2578153.2578168

1 Introduction

Duchowski et al. [2011; 2012] have reported an increase in mean signed and squared vergence error with distance from the screen when measuring eye-tracked gaze depth while viewing stereoscopic displays. Their analysis of signed error (based on gaze disparity measured at the screen) shows a bias toward underestimation of target visual distance both in front of and behind the screen. Such findings may be relevant to the breadth of research documenting underestimation of reaching distance in similar (virtual) environments [Napieralski et al. 2011; Kellner et al. 2012]. Distance estimation in reaching tasks when looking at stereo objects (touching objects displayed with positive or negative stereoscopic parallax, i.e., disparity, on a 2D surface) induces more imprecision than touching objects with zero parallax [Valkov et al. 2011].

Vergence error may be biased toward the fixed accommodative distance of the display, however, it is also a function of the depth estimate itself, which is derived relative to the 2D calibration plane of the eye tracker. In this paper we compare gaze measured binocularly when viewing a haploscope in virtual reality to gaze measured binocularly when performing a similar task in physical reality. Along with comparison of vergence error at different depths, we detail techniques developed for gaze data analysis using a first-of-its kind, commercially available binocular head-mounted eye tracker.

We make three contributions related to gaze depth measurements in the physical environment: (1) the distance to the 2D calibration plane is critical as it bears on subsequent estimation of gaze depth; (2) although highly correlated with each other, a vector-based gaze depth estimate appears more accurate than one based on disparity as the latter appears to inflate error magnitude; (3) contrary to the virtual stereoscopic display, the magnitude of vergence error in physical reality appears to increase with distance from the viewer.

2 Previous Work

Daugherty et al. [2010] were one of the first to use a commodity desktop binocular eye tracker to measure gaze vergence. Earlier efforts also measured vergence, but they used more specialized equipment [Essig et al. 2004; Medlin 2003; Duchowski et al. 2002]. More recently, using commodity eye trackers, Wang et al. [2012] provided details regarding estimation of gaze depth along with associated signal filtering and 3D eye tracker calibration techniques.

Gaze depth, i.e., vergence, can be measured via *spatial triangulation* [Pfeiffer 2010]. The geometry of the approach is given by Wang et al. [2013]. If a binocular eye tracker delivers two on-screen gaze points, (x_l, y_l) for the left eye and (x_r, y_r) for the right, then the horizontal disparity $\Delta x = x_r - x_l$ is sufficient to estimate gaze depth z [Duchowski et al. 2011]. For a review of eye-tracked 3D gaze estimation, see Pfeiffer [2012] and Essig et al. [2012].

In this paper, we compare 3D gaze estimation via triangulation in two environments: a screen-based haploscope constituting a stereoscopic virtual environment display and a screen-less environment composed of physical (planar) targets. Gaze depth in the physical environment is, to our knowledge for the first time, computed via spatial triangulation from two separate gaze data streams, each providing (x_l, y_l) and (x_r, y_r) gaze points mapped from two separate (head-mounted) eye cameras to a common scene camera reference frame. The homography is computed by the eye tracker vendor's proprietary software. We describe the left and right gaze stream synchronization with computation of gaze depth, and report on gaze depth error computed to physical targets identified in the scene camera. We compare the error to that measured on the haploscope with a similarly arranged virtual scene.

3 Technical Development

We used two eye-tracking systems to binocularly measure gaze from which we computed a gaze depth (vergence) estimate: a stereoscopic display projecting a virtual scene; and a physical scene with targets arranged to resemble those shown virtually.

3.1 Apparatus

The virtual environment was displayed on a Wheatstone-style haploscope, first reported by Bair et al. [2006], shown in use in Figure 1(a). It consists of two high-resolution *IBM T221* "Big Bertha" LCD monitors driven by two *NVidia Quadro FX 5800* graphics cards installed in an *Intel Xeon E5502 Dualcore* PC running the *CentOS* operating system. Both displays are set on a track, sitting 172 cm apart on opposite sides of two small mirrors angled at 45° from the medial axis. The screens are 48 cm wide and 30 cm high (16:10), with a screen resolution of 3840×2400, or 9.2 million pixels. The screens are set at a viewing distance of 86.36 cm from the nominal eye point. The visual angle subtended by one pixel is about 0.5 arc minutes, corresponding closely to the foveal resolution of the eye [Campbell and Green 1965]. For stereoscopic display, images were rendered assuming an eye separation of 6.3 cm, with a viewing frustum corresponding to the physical setup. The monitors and mirrors are carefully aligned so that when looking into the mirrors, the images on both monitors are fused into a single virtual image.

Eye tracking cameras (monocular) from *LC Technologies* are mounted beneath each monitor, part of the *Eyegaze System* that is used to image the viewer's eyes as seen by the cameras in the mirrors. According to the manufacturer's brochure, each camera operates at a 60 Hz sampling rate with an accuracy of $< 0.4°$.

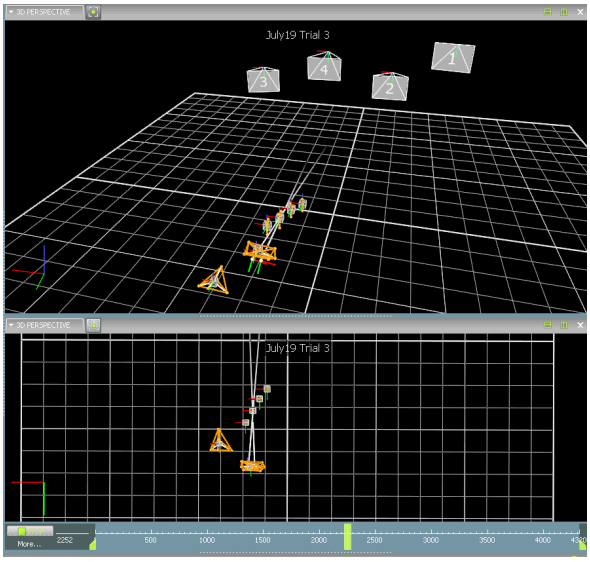


Figure 2: Vicon*'s physical environment model (c.f. Figure 1(b)).*

The physical environment was constructed by placing four visual targets on a flat surface in front of the viewer (see § 4), shown in Figure 1(b). The eye tracker used in this environment is the *Dikablis* from *Ergoneers*. The unit is head-mounted, custom-built with binocular cameras, one pointing at each eye from below. Each eye camera operates at 25 Hz, and according to the manufacturer, measures gaze direction with an accuracy $< 0.5°$. A third camera faces forward, capturing the scene in front of the viewer.

The *Dikablis* eye tracker is meant to interface with *Vicon*'s motion capture system so that gaze direction can be retrieved in a model of the 3D world captured by *Vicon*'s cameras. Four *Vicon* T40-S cameras were placed 320 cm in front of and facing the viewer. According to the manufacturer, each of the T40-S cameras has resolution of 4 megapixels, and functions at maximum frame rate of 515 frames per second. *Vicon*'s cameras capture the motion of objects in their field of view when adorned with infra-red reflective markers. The *Dikablis* headband is outfitted with plastic "antlers" housing 6 such markers (3 on each side) that sit roughly at one's forehead, which provide position and orientation information relative to *Vicon*'s 3D world origin. Each of the visual targets used in the experiment was also adorned with three such infra-red reflective markers, affixed to its back surface that faced the cameras. Although *Vicon*'s *Tracker* (v2.0) software was used to record 3D world and gaze data captured in the experiment (see Figure 2), in this paper we discuss gaze depth estimation which we computed only from the data recorded by *Dikablis*.

3.2 Vergence (Depth) Estimation from Gaze Disparity

Gaze depth z is derived relative to the screen, taking the screen center as the origin, and assuming that the viewing position is aligned with the screen center. The viewing distance D to the screen is measured and set dependent on the apparatus used. Inter-ocular distance α can be assumed to be constant at 6.3 cm [Smith and Atchison 1997]. Although eye separation varies across individuals [Dodgson 2004], and significantly affects spatial triangulation, for a fixed-screen display a 3D calibration procedure is used to accommodate for this variation [Wang et al. 2013] (see below).

Given gaze disparity $\Delta x = x_r - x_l$, the disparity induced gaze

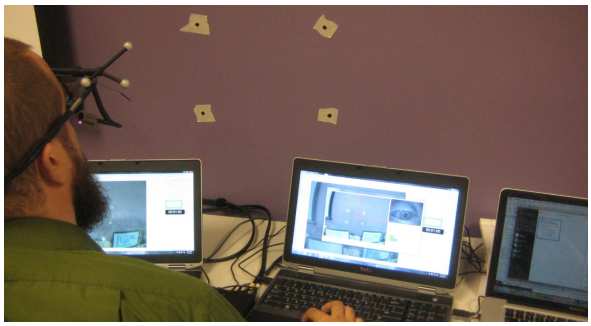


Figure 3: *2D calibration of the* Dikablis *system.*

depth z in screen-centered coordinates is given by

$$\frac{-z}{\Delta x} = \frac{D-z}{\alpha} \quad \Rightarrow \quad z = \frac{\Delta x D}{\Delta x - \alpha}, \tag{1}$$

where $z = 0$ denotes gaze depth at the screen plane, with z positive in front of the screen, and negative behind.

3.3 2D and 3D Calibration on the *Wheatstone*

Prior to commencing experimental trials on the haploscope, each participant first completed a 13-point 2D calibration, which is standard on the eye tracking vendor's system. 2D calibration produces 2D gaze coordinates (x_l, y_l) and (x_r, y_r) which are then used to estimate gaze depth z with Eq. (1).

Wang et al. [2012] showed that estimated gaze depth can be noisy, and prone to estimation errors. To reduce these effects, they introduced online filtering followed by a first order online calibration for smoothing and adjusting gaze depth. We implement their filtering and 3D calibration, which follows successful completion of 2D calibration. Each viewer is asked to visually pursue a calibration point (a simple sphere) translating through space along a Lissajous-knot path. Collection of calibration data takes 40 seconds, during which the sphere's position $p(t) = (x(t), y(t), z(t))$ changes with time t in seconds, according to $p(t) = \mathbf{A}\cos(2\pi\,\mathbf{f}\,t + \phi)$, with component amplitude $\mathbf{A} = (9, 5, 20)$ cm, frequencies $\mathbf{f} = (0.101, 0.127, 0.032)$ Hz, and phase angles $\phi = (0°, -90°, 57°)$. Letting $\mathbf{S} = \begin{bmatrix} s_{iz} \end{bmatrix}$ denote the known depth coordinates of the calibration sphere, z_i the measured depth coordinates computed by Eq. (1) from the filtered disparity measurements at each time sample i, and letting $\mathbf{B} = \begin{bmatrix} b_0 & b_1 \end{bmatrix}$ be unknown coefficients, a first order model $\mathbf{S} = \begin{bmatrix} 1 & z_i \end{bmatrix}\mathbf{B}$ is used to effect calibration, which minimizes errors due to systematic shift (from $z = 0$) and depth-scale errors. The general solution for estimation of $\mathbf{B}$ relies on Lagrange's method of least squares, or the multivariate multiple regression model (see Wang et al. [2012; 2013] for details).

Online filtering is performed prior to 3D calibration by first cleaning horizontal disparity Δx via an online outlier removal process. This is done by replacing values beyond two standard deviations with the most recent valid reading. Common running mean and running standard deviation computations were used [Brown 1983]. Outlier removal is followed by filtering through an online implementation of a 6^{th} order Butterworth filter, whose cutoff frequency was set at 0.15 Hz, against a tracker sampling frequency of 60 Hz (see Duchowski et al. [2011] for details).

3.4 2D and 3D Calibration on the *Dikablis*

The binocular *Dikablis* system streams monocular gaze estimates for the each of the left and right eyes, as well as the corresponding

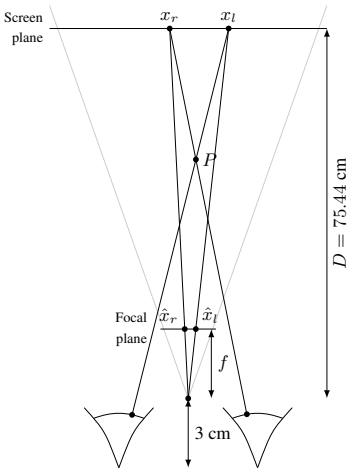


Figure 5: Dikablis *gaze estimation.*

eye and scene videos to two separate laptops (see Figure 3). Although the data is timestamped by each laptop, we could not guarantee that the systems clocks of these computers were well synchronized. Instead, since the scene videos on both laptops originated from the same camera, we could therefore align the gaze data by aligning the videos. As the experiments were short, we assumed that clock skew was negligible, and we only needed to determine the clock offset. We found this offset by minimizing the sum of square per-pixel differences between aligned left-scene and right-scene video frames.

The synchronized *Dikablis* data provides a left and right gaze point in the scene video (see Figure 4(a)). As there is no physical screen on which the tracking is being performed, the relationship of these gaze points to the 3D gaze location is determined by the initial 2D calibration. The *Dikablis* 2D calibration procedure, performed once for each of the left and right eyes, requires the user to look at four points on a plane—in our case, points on a wall (see Figure 3). In effect, this calibration plane then becomes the screen plane, and so gaze points are the intersections of the gaze vectors with the calibration plane, projected onto the scene camera's image plane.

Estimating gaze depth using Eq. (1) requires Δx, the gaze disparity on the screen plane. The raw *Dikablis* data only gives projected gaze points in pixels, so we first unproject these gaze points onto the screen plane. The unprojected coordinates (x_l, y_l) and (x_r, y_r) are calculated from the raw coordinates $(\hat{x}_l, \hat{y}_l)$ and $(\hat{x}_r, \hat{y}_r)$ as:

$$\begin{aligned} x_l &= \hat{x}_l D/f \quad & x_r &= \hat{x}_r D/f \\ y_l &= \hat{y}_l D/f \quad & y_r &= \hat{y}_r D/f \end{aligned} \tag{2}$$

where $D = 75.44$ cm is the unprojection distance (in our case, to the wall) and $f = 484.23$ pixels is the focal length of the camera.

We found the focal length of the camera using the standard chessboard camera calibration introduced by Zhang [2000] and implemented in `OpenCV`. We also used the same method to find the distortion parameters of the camera, and used these to correct for the spherical distortion of the *Dikablis* scene camera (see Figure 4).

Screen distance was obtained by placing a marker of known size on the calibration wall, and estimating its 3D position using `OpenCV`'s Levenberg-Marquardt-based Perspective-n-Point (PnP) solver.

Given the focal length and screen distance, we could then calculate the left and right screen gaze points using Eq. (2), and calculate gaze depth from their disparity using Eq. (1). Note that for gaze

Figure 6: *3D calibration of the* Dikablis *system.*

depth calculation we used $D = 78.44$ cm to account for an estimated offset from the scene camera (see Figure 5).

We also used a direct gaze vector intersection method to calculate a second estimate of the gaze depth. We used the left and right gaze vectors—that is, vectors passing through the corresponding screen gaze points and eye centers—to compute the intersection of line-of-sight in 3D space (see Figure 5). Because line-of-sight intersection is unlikely, gaze depth is estimated as the midpoint of the shortest line between the gaze vectors [Wibirama and Hamamoto 2013]. We used the z-coordinate of this gaze point P as the gaze depth $\vec{z}$.

Dikablis does not provide a 3D calibration and so we did not apply a type of linear correction to the disparity depth estimate z as we did for the data recorded on the Wheatstone tracker. However, a 3D calibration is provided by *Vicon*. Although we did not use it in our current analysis, we did collect this data for future investigation. We quickly review *Vicon*'s 3D calibration procedure for completeness in describing the use of this novel system.

Following 2D calibration for each of the left and right eyes, while the two laptops save the monocular gaze estimate streams to disk (which is only what we used in the present analysis), they also stream the data to *Vicon*'s *Tracker* software. Once a TCP/IP connection is established to the *Dikablis* laptops, both left and right gaze streams are available to *Tracker* as eye tracker objects. These in turn are added and their positions are set relative to the head object tracked by *Tracker* (the IR-reflective "antlers" worn by the user). We used offsets recommended by *Vicon*, namely $\{30, 40, -40\}$ mm for the $\{x, y, z\}$ left eye offsets and $\{-30, 40, -40\}$ mm for the right. Once this is accomplished, 3D calibration can be performed. To do so, for each point added, *Vicon*'s wand is held at an arbitrary distance from the user, as shown in Figure 6. Following 3D calibration, gaze vectors are visible emanating from the eyeball objects, as seen in Figure 2.

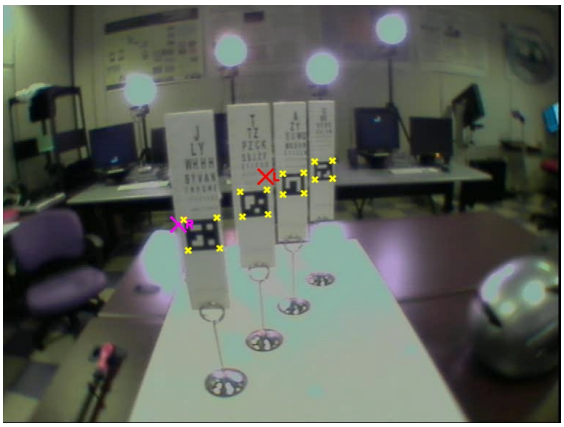
To partially match the signal processing used in the Wheatstone apparatus, gaze depths z and $\vec{z}$ collected with *Dikablis* were filtered (offline) with a 6^{th} order Butterworth filter, with cutoff and sampling frequencies set to 0.15 Hz and 25 Hz, respectively.

4 Methodology

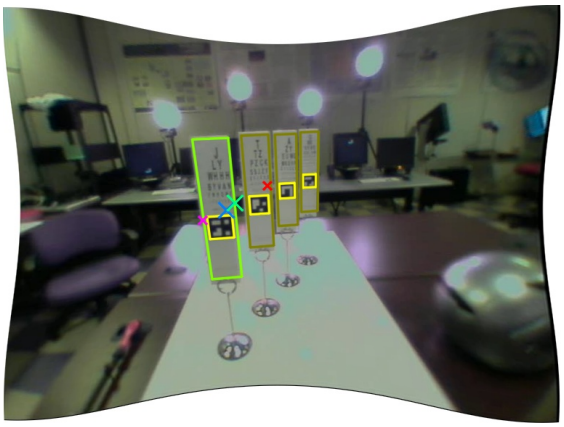
Participants viewed a simple scene in both virtual and physical environments, which was designed to appear similarly in both, with particular attention paid to the relative depths of the visual targets.

Stimulus. The virtual environment consisted of a scene with four Snellen charts, patterned after the stimulus used by Love et al. [2009], but arranged so that no chart was occluded by another (see Figure 7). The targets are arranged at four depths relative to the ($z = 0$) screen (from back-to-front): -15, -5, 5, 15 cm, with negative distances behind the screen. Their horizontal offsets from the screen center are (from back-to-front): 9.5, 2.5, -3.5, and -9.5 cm, with negative distances left of screen center.

Because stereoscopic displays dissociate the natural coupling between vergence and focal distance (accommodation) by rendering images with non-zero disparity (stimulating vergence) at a fixed display distance [Wann et al. 1995; Wilson 1996; Rushton and Riddell 1999], they have been considered to cause visual discomfort, with its source tied to eye strain and fatigue [Howard and Rogers 2002; Iwasaki et al. 2009]. We decided to test the efficacy of gaze-contingent depth-of-field used to blur regions of the display outside the viewer's eye-tracked gaze depth [Mantiuk et al. 2011]. Unlike Mantiuk et al., we do not set the depth-of-field focal plane to the depth value of the current pixel at the viewer's gaze point coordinates, rather, we measure depth-of-field focal plane to the estimated gaze depth z. We were particularly interested in finding whether

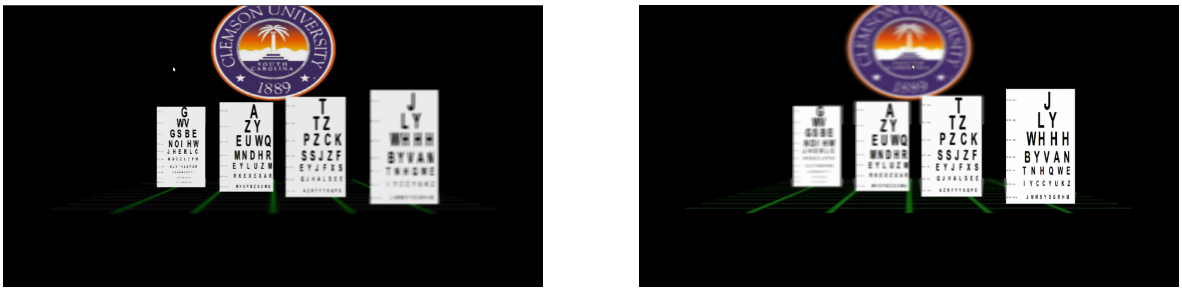


(a) Synchronized Dikablis *raw data, including left and right gaze points, with tracked marker corners.*



(b) Processed Dikablis *data, including gaze point calculation via disparity (green cross) and gaze vector intersection (blue cross), marker pose (yellow squares), and current detected gaze target (green rectangle).*

Figure 4: Dikablis *data processing.*



(a) Looking at far chart. *(b) Looking at near chart.*

Figure 7: *Snellen eye chart images used in the experiment, shown here in the virtual environment with depth-of-field effect.*

this attenuation of peripheral accommodative demand would have an effect on vergence error in the stereoscopic environment.

The physical environment was constructed by printing out the four Snellen charts, affixing them to cardboard, and mounting them on 6 inch metallic stands fixed on a flat surface, 6 cm apart side-to-side, and 10 cm apart in depth. Their approximate distances to viewers was (from back-to-front): 70, 60, 50, and 40 cm. Taking the *Dikablis* calibration plane distance (75.44 cm) into account, the eye charts were approximately at the following depths relative to the calibration plane (from back-to-front): 5, 15, 25, 35, so that all were effectively positioned in front of the calibration plane.

Experimental Design. We were mainly concerned with the within-subjects comparison of vergence error within the two environments. We were also concerned with order effects, i.e., viewing of the eye charts as well as of exposure to each environment. We therefore conducted a $2\times2\times2\times2$ mixed-design experiment. Our main within-subjects factor was the apparatus at 2 levels (virtual or physical). To control for order effects, we randomly assigned half the participants to start in the virtual environment, the other in the physical, thus constituting the second between-subjects factor of apparatus order. Similarly, we controlled for reading order where half the participants read the charts left-to-right, the other half right-to-left, giving the third between-subjects factor. Finally, we considered gaze-contingent depth-of-field as the fourth between-subjects factor (experienced by half the participants).

We assigned each participant to one of the four following conditions: (A) *Wheatstone* first, then *Dikablis*, left-to-right viewing; (B) *Wheatstone* first, then *Dikablis*, right-to-left viewing; (C) *Dikablis* first, then *Wheatstone*, left-to-right viewing; (D) *Dikablis* first, then *Wheatstone*, left-to-right viewing. Half the participants saw the depth-of-field effect, half did not.

Participants. Twenty-one participants took part in the study (10 M, 11 F) with data from two male participants excluded from the analysis due to procedural problems. Of those included in the analysis, 9 participants were exposed to the gaze-contingent depth-of-field effect (5 M, 4 F), and 10 participants were not (3 M 7 F); all 19 thus took part in both virtual and physical environments. For analysis, data from 4 participants were removed either due to their outlier characteristics with respect to the main dependent measures z or $\vec{z}$,[1] or due to missing data caused by procedural errors.

[1]Outliers are defined above the 3^{rd} or below the 1^{st} quartile $\pm$ 1.5 $\times$ the inner quartile range, respecitvely, as per the standard method used by R's boxplot function.

5 Results

In the virtual environment, vergence error is determined by ray casting the left and right gaze direction bisector. Error is computed as the signed distance between intersected target depth and gaze depth. Examples of detected target and gaze depth are shown in Figure 8.

In the physical environment, targets were identified by visual (fiducial) markers attached to the visual stimulus (fiducial markers were situated beneath the Snellen charts). To calculate gaze error, we used the 3D positions of these markers as derived from their 2D positions provided by *Dikablis* marker tracking software.

Specifically, *Dikablis* raw data includes 2D marker tracking, which gives the positions of the four corners of each marker in the scene video. We used these corners and known marker size to estimate the 3D position of each marker individually, again using `OpenCV`'s PnP solver. To obtain a more robust estimate of marker position, we wanted to track the entire marker assembly rather than the individual markers. We first calculated the markers' fixed positions relative to each other by averaging the individual marker positions over the course of the video. We then used these relative positions as input into the PnP solver to obtain a pose estimate of the entire marker assembly which included the targets fixated by participants.

To detect which marker a user looked at, we extended the 3D marker model to include the card on which the marker was printed. For each marker, we reprojected the 3D card rectangle and the gaze point onto the scene camera's focal plane, and performed a simple point-in-polygon test to determine if the projected gaze point was within the projected marker card polygon. If it was, we assumed that the user was looking at that marker. See Figure 9 for visualization of the physical model.

5.1 Vergence Errors in the Physical Environment

Pearson's correlation analysis of the disparity-based depth estimate z to the vector intersection-based estimate $\vec{z}$ revealed a highly positive significant coefficient, $R = 0.77, t(54) = 8.74, p < 0.01$, suggesting that both methods of error calculation give very similar relative results across depths. Moreover the correlation coefficient for z and $\vec{z}$ at different target depths increases systematically: $R(12) = 0.59$ at 5 cm, to $R(12) = 0.77$ at 15 cm, to $R(12) = 0.82$ at 25 cm, to $R(12) = 0.90$ at 35 cm (see Figure 10). Fisher's r-to-z transformation was used to test for statistical significance between correlation coefficients. No significant differences were detected.

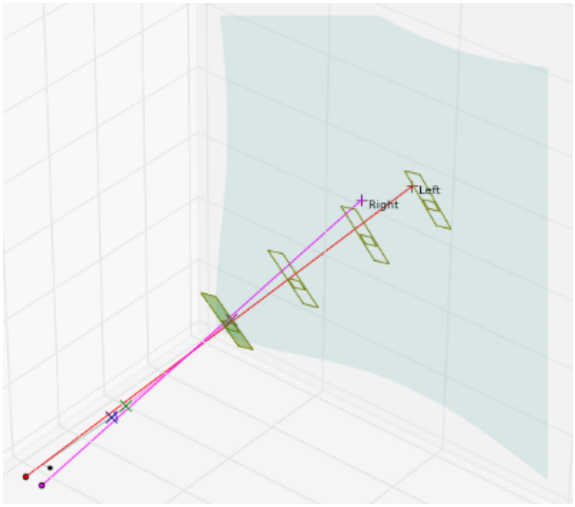


Figure 9: *A 3D visualization of the processed Dikablis data. The red and pink lines are the gaze lines, the yellow rectangles are the markers and the blue surface is the screen plane.*

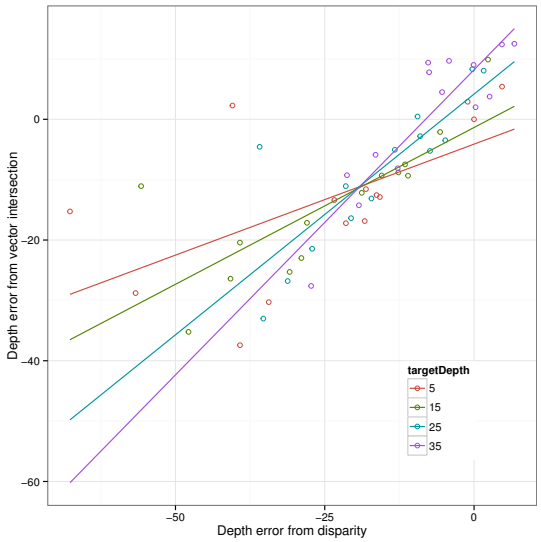


Figure 10: *Correlations between z and $\vec{z}$ at target depths.*

5.2 Vergence Errors: Physical vs. Virtual Environment

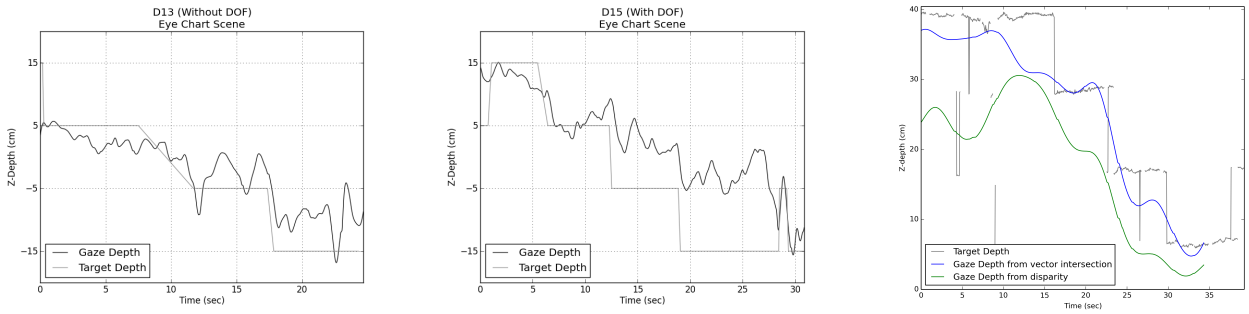
The influence of target depth on vergence error was examined via a series of one-sample t tests conducted to gauge whether vergence errors differed significantly from the expectation of no error (zero error). Analyses were performed separately in the virtual and physical environments, with separate evaluation of z and $\vec{z}$ in the latter.

Results from measurements obtained on the *Wheatstone* suggest greater signed vergence error with negative conflict (stereo conflict behind the screen) than positive conflict (stereo in front of the screen). Vergence errors ($M = 7.76, SE = 2.42$) at -15 and 15 cm depths ($M = -6.66, SE = 2.41$) were significantly different from zero, while differences at -5 ($M = 4.22, SE = 2.61$) and 5 cm ($M = 1.8, SE = 2.19$) were not (see Figure 11 and Table 1). This is likely because small errors in gaze disparity (convergence or divergence) lead to large errors in the vergence estimate.

A one-way within-subjects ANOVA was used to check for significant differences in vergence errors at different depths in the virtual environment. The main effect of target depth was significant, $F(3, 33) = 12.22, p < 0.01$. The following pairwise comparisons with Bonferroni correction revealed significantly different errors between depths of 5 and 15, -5 and 15, and between 15 and -15 cm. Results generally corroborate previous findings of signed vergence error on stereoscopic displays [Wang et al. 2013].

Similar analyses were conducted on the *Dikablis* for both errors z and $\vec{z}$. For error z, all were significantly smaller than zero at all depths (see Table 1 and Figure 11). The following one-way within-subjects analysis of variance comparing mean errors between all depths showed that they differed significantly, $F(3, 39) = 10.11, p < 0.01$. Pairwise comparisons with Bonferroni correction revealed significant differences ($p < 0.01$) between depths of 5 ($M = -24.86, SE = 5.60$) and 35 cm ($M = -7.69, SE = 2.79$), as well as between 35 and 15 cm ($M = -24.59, SE = 4.54$), and between 35 and 25 cm ($M = -16.53, SE = 3.34$). The remaining differences were not significant.

According to t tests, vergence error $\vec{z}$ is significantly different from 0 at depths of 5, 15, and 25 cm, but not at 35 cm (see Table 1 and Figure 11). Analysis of variance revealed a significant main effect of target depth, $F(3, 39) = 8.88, p < 0.01$. Pairwise comparisons with Bonferroni correction showed that $\vec{z}$ errors differ significantly between targets at 5 ($M = -13.26, SE = 3.46$) and 35 cm ($M = 0.42, SE = 3.16$), 15 ($M = -14.13, SE = 3.06$) and 35 cm, and 25 ($M = -9.01, SE = 3.25$) and 35 cm, and 15 cm vs. 25 cm.



(a) Viewing right-to-left without depth-of-field. *(b) Viewing right-to-left with depth-of-field.* *(c) Viewing right-to-left.*

Figure 8: *Vergence error computed on: (a) & (b)* Wheatstone*; and (c)* Dikablis*.*

Depth	*Wheatstone*	*Dikablis z*	*Dikablis $\vec{z}$*
-15	$t(11)=3.21, p<0.01$		
-5	$t(11)=1.61$, n.s.		
5	$t(11)=0.82$, n.s.	$t(13)=-4.44, p<0.01$	$t(13)=-3.83, p<0.01$
15	$t(11)=-2.76, p<0.02$	$t(13)=-5.42, p<0.01$	$t(13)=-4.62, p<0.01$
25		$t(13)=-4.95, p<0.01$	$t(13)=-2.77, p<0.02$
35		$t(13)=-2.76, p<0.02$	$t(13)=0.13, n.s.$

Table 1: *Vergence error comparisons to expectation of no (zero) error at different target depths.*

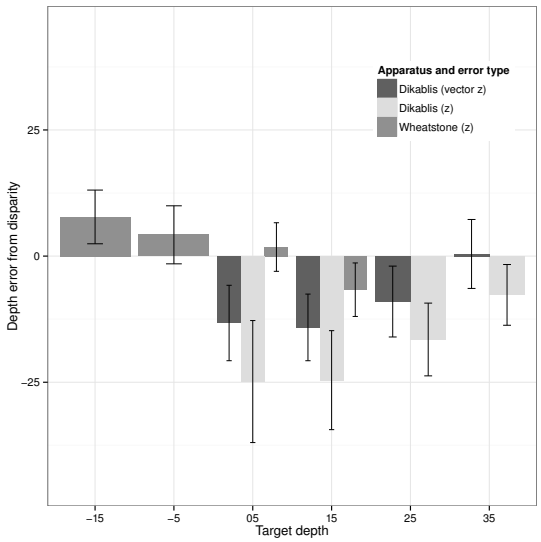


Figure 11: *Vergence error at target depths, relative to screen plane.*

On average, comparing vergence error between apparatus types, combining errors z and $\vec{z}$ on the *Dikablis*, a one-way within-subjects ANOVA with apparatus treated as a fixed factor reveals a significant main effect of apparatus $F(1,13) = 14.01, p < 0.01$, showing that signed vergence error on the *Dikablis* was significantly greater ($M = -8.99, SE = 1.76$) than on the *Wheatstone* ($M = 1.54, SE = 1.33$), when considering error magnitude.

5.3 Effects of Reading and Apparatus Orders

Between-subjects ANOVAs were conducted to test the influence of reading order on vergence error, revealing the main effect of reading order was not significant, $F(1,10) < 1$. The influence of apparatus order on vergence error was tested via between-subjects one-way ANOVAs with reading order as the independent variable. Results showed a main effect of apparatus order at a statistical tendency level, $F(1,12) = 4.63, p = 0.058$, with signed vergence error generally greater ($M = -0.16, SE = 3.21$) when the *Wheatstone* was used ahead of the *Dikablis* ($M = -6.55, SE = 3.35$).

5.4 Effect of Depth-Of-Field on Vergence Error

To test the influence of depth-of-field (DOF) on vergence error, a two-way mixed-design 2×4 ANOVA was conducted. In this analysis target depth was treated as a within-subjects factor and depth-of-field a between-subjects factor. The dependent variable was the magnitude of error based on disparity. For the analysis we have used only data from the *Wheatstone* since the manipulation of DOF was performed only on this apparatus. The main effect of DOF reached significance, $F(1,10) = 8.07, p < 0.02$, with the magnitude of vergence error significantly lower with depth-of-field absent ($M = -3.13, SE = 1.57$) than when present ($M = 5.29, SE = 1.88$). Analysis also revealed a significant effect of target depth, $F(3,30) = 11.52, p < 0.01$, consistent with previous analyses (see above). The interaction effect of DOF and target depth was not statistically significant, $F(3,30) = 1.77, p > 0.1$.

6 Discussion, Limitations, and Future Work

Vergence error estimates in the physical environment (z and $\vec{z}$) correlate better as target depth increases from the 2D calibration plane, i.e., as target depth decreases from the viewer. As negative disparity i.e., when $x_r - x_l < 0$, increases (due to convergence in this case), the estimate of gaze depth improves. As in the virtual environment, small errors in gaze disparity lead to large errors in the estimate.

Results suggest that vergence error is relatively larger when viewing the physical setup (in front of the calibration plane) than when viewing the stereoscopic display (when objects are projected in front of the screen). In the physical setup, where there is no accommodation-vergence conflict, vergence error is greatest at the target that is closest to the calibration plane (at $z = 5$) but farthest away from the viewer (at approximately 70 cm; see Figure 11).

Why should vergence error in the physical environment increase (away from the viewer) toward the screen plane? In the absence of a display screen, perhaps vergence mimics underestimation of reaching response observed by Napieralski et al. [2011]. In *near space* (distance up to an arm's length), underestimation of physical reach responses to targets increases according to a linear regression model with slope 0.77 and intercept 0.02. For example, at half arm's distance, reach responses underestimated target distance by $0.77 \times .5 + 0.02 = 0.41$ meaning the distance to target is underestimated by 9%. Meanwhile, at full arm's distance, target distance is underestimated by 21%. This makes sense intuitively as one is likely to make better distance judgments at closer distances [Cutting and Vishton 1995]. Although this was not expected at the outset, vergence estimates may reflect this in near space and may be more error-prone at farther distances. When looking at a display, however, vergence may be mediated by the screen distance.

A more careful comparison of vergence at similar target depths is needed. The initial goal of the study was to compare vergence error in both physical and virtual environments at comparable target depths. Physical targets were set in near space because we were concerned about the *Dikablis* system being able to locate the fiducial markers used to identify visual targets. We limited the distance of both the farthest physical eye chart and the calibration plane to just beyond arm's reach. At the outset, the importance of the distance to the calibration plane was not evident. Misalignment of the physical targets to their virtual counterparts is clearly a limitation of the present study. Future experiments comparing vergence as computed from binocular disparity must carefully consider the relative distances to the planes at which initial gaze disparity is calculated.

7 Conclusion

The disparity-based estimate of gaze depth, while seemingly appropriate in stereoscopic virtual reality, appears to inflate vergence error to targets in physical reality. Although distance to the 2D calibration plane in both environments affects both estimates, when a physical fixed accommodation screen is absent, vergence computation should rely on the vector-intersection model.

References

BAIR, A., HOUSE, D., AND WARE, C. 2006. Texturing of Layered Surfaces for Optimal Viewing. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Visualization) 12*, 5, 1125–1132.

BROWN, R. G. 1983. *Introduction to Random Signal Analysis and Kalman Filtering*. John Wiley & Sons, New York, NY.

CAMPBELL, F., AND GREEN, D. 1965. Monocular versus binocular visual acuity. *Nature 209*, 191–192.

CUTTING, J. E., AND VISHTON, P. M. 1995. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of Space and Motion*, W. Epstein and S. Rogers, Eds. Academic Press, San Diego, CA, 69–117. (Handbook of Perception and Cognition, Vol. 5).

DAUGHERTY, B. C., DUCHOWSKI, A. T., HOUSE, D. H., AND RAMASAMY, C. 2010. Measuring Vergence Over Stereoscopic Video with a Remote Eye Tracker. In *ETRA '10: Proceedings of the 2010 Symposium on Eye Tracking Research & Applications (Late Breaking Results)*, ACM, Austin, TX.

DODGSON, N. A. 2004. Variation and extrema of human interpupillary distance. In *Stereoscopic Displays and Virtual Reality Systems XI*, A. J. Woods, J. O. Merritt, S. A. Benton, and M. T. Bolas, Eds., vol. 5291, SPIE, 36–46.

DUCHOWSKI, A., MEDLIN, E., COURNIA, N., GRAMOPADHYE, A., MELLOY, B., AND NAIR, S. 2002. 3D Eye Movement Analysis for VR Visual Inspection Training. In *ETRA '02: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ACM, New Orleans, LA, 103–110,155.

DUCHOWSKI, A. T., PELFREY, B., HOUSE, D. H., AND WANG, R. 2011. Measuring Gaze Depth with an Eye Tracker During Stereoscopic Display. In *Applied Perception in Graphics & Visualization (APGV)*, ACM.

DUCHOWSKI, A. T., PELFREY, B., HOUSE, D. H., AND WANG, R. 2012. Eye-Tracked Vergence Response During Active-Stereo Display. In *The 5th Hamlyn Symposium on Medical Robotics*.

ESSIG, K., POMPLUN, M., AND RITTER, H. 2004. Application of a Novel Neural Approach to 3D Gaze Tracking: Vergence Eye-Movements in Autostereograms. In *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*, K. Forbus, D. Gentner, and T. Regier, Eds., Cognitive Science Society, 357–362.

ESSIG, K., PRINZHORN, D., MAYCOCK, J., DORNBUSCH, D., RITTER, H., AND SCHACK, T. 2012. Automatic analysis of 3d gaze coordinates on scene objects using data from eye tracking and motion-capture systems. In *ETRA '12: Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY.

HOWARD, I. P., AND ROGERS, B. J. 2002. *Seeing in Depth*, vol. II: Depth Perception. I Porteous, University of Toronto Press, Thornhill, ON, Canada.

IWASAKI, T., KUBOTA, T., AND TAWARA, A. 2009. The tolerance range of binocular disparity on a 3D display based on the physiological characteristics of ocular accommodation. *Displays 30*, 44–48.

KELLNER, F., BOLTE, B., BRUDER, G., RAUTENBERG, U., STEINICKE, F., LAPPE, M., AND KOCH, R. 2012. Geometric Calibration of Head-Mounted Displays and its Effects on Distance Estimation. *IEEE Transactions on Visualization and Computer Graphics 18*, 4, 589–596.

LOVE, G. D., HOFFMAN, D. M., HANDS, P. J. W., GAO, J., KIRBY, A. K., AND BANKS, M. S. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express 17*, 18, 15716–15725.

MANTIUK, R., BAZYLUK, B., AND TOMASZEWSKA, A. 2011. Gaze-Dependent Depth-of-Field Effect Rendering in Virtual Environments. In *SGDA '11: Serious Games Development and Applications*, Springer-Verlag, 1–12.

MEDLIN, E. 2003. *Ocular Vergence Measurement in Bi[n]ocular Virtual Reality*. Master's thesis, Clemson University, Clemson, SC.

NAPIERALSKI, P. E., ALTENHOFF, B. M., BERTRAND, J. W., LONG, L. O., BABU, S. V., PAGANO, C. C., KERN, J., AND DAVIS, T. A. 2011. Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Transactions on Applied Perception 8*, 3, 18:1–18:19.

PFEIFFER, T. 2010. *Understanding Multimodal Deixis with Gaze and Gesture in Conversational Interfaces*. PhD thesis, Bielefeld University, Bielefeld, Germany.

PFEIFFER, T. 2012. Measuring and visualizing attention in space with 3d attention volumes. In *ETRA '12: Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY.

RUSHTON, S. K., AND RIDDELL, P. M. 1999. Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence. *Applied Ergonomics 30*, 69–78.

SMITH, G., AND ATCHISON, D. A. 1997. *The Eye and Visual Optical Instruments*. Cambridge Univ. Press, Cambridge, UK.

VALKOV, D., STEINICKE, F., BRUDER, G., AND HINRICHS, K. 2011. 2D Touching of 3D Stereoscopic Objects. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, 1353–1362.

WANG, R., PELFREY, B., DUCHOWSKI, A. T., AND HOUSE, D. H. 2012. Online Gaze Disparity via Binocular Eye Tracking on Stereoscopic Displays. In *Second Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization & Transmission (3DimPVT 2012)*, IEEE.

WANG, R. I., PELFREY, B., DUCHOWSKI, A. T., AND HOUSE, D. H. 2013. Online 3D Gaze Localization on Stereoscopic Displays. *Transactions on Applied Perception*. (In press.).

WANN, J. P., RUSHTON, S., AND MON-WILLIAMS. 1995. Natural Problems for Stereoscopic Depth Perception in Virtual Environments. *Vision Research 35*, 19, 2731–2746.

WIBIRAMA, S., AND HAMAMOTO, K. 2013. 3D Gaze Tracking System for NVidia 3D Vision. In *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBS)*.

WILSON, J. R. 1996. Effects of Participating in Virtual Environments: A Review of Current Knowledge. *Safety Science 23*, 1, 39–51.

ZHANG, Z. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 11 (Nov.), 1330–1334.