Vergence–accommodation conflicts hinder visual performance and cause visual fatigue

David M. Hoffman,
Vision Science Program, University of California, Berkeley, CA, USA

Ahna R. Girshick,
Vision Science Program, University of California, Berkeley, CA, USA

Kurt Akeley, and
Microsoft Research Silicon Valley, Mountain View, CA, USA

Martin S. Banks
Vision Science Program, Department of Psychology Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

Abstract

Three-dimensional (3D) displays have become important for many applications including vision research, operation of remote devices, medical imaging, surgical training, scientific visualization, virtual prototyping, and more. In many of these applications, it is important for the graphic image to create a faithful impression of the 3D structure of the portrayed object or scene. Unfortunately, 3D displays often yield distortions in perceived 3D structure compared with the percepts of the real scenes the displays depict. A likely cause of such distortions is the fact that computer displays present images on one surface. Thus, focus cues—accommodation and blur in the retinal image—specify the depth of the display rather than the depths in the depicted scene. Additionally, the uncoupling of vergence and accommodation required by 3D displays frequently reduces one's ability to fuse the binocular stimulus and causes discomfort and fatigue for the viewer. We have developed a novel 3D display that presents focus cues that are correct or nearly correct for the depicted scene. We used this display to evaluate the influence of focus cues on perceptual distortions, fusion failures, and fatigue. We show that when focus cues are correct or nearly correct, (1) the time required to identify a stereoscopic stimulus is reduced, (2) stereoacuity in a time-limited task is increased, (3) distortions in perceived depth are reduced, and (4) viewer fatigue and discomfort are reduced. We discuss the implications of this work for vision research and the design and use of displays.

Keywords

accommodation; vergence; volumetric display; stereopsis; binocular vision; 3D display; visual fatigue; cue combination; depth perception; graphics

Introduction

Consider two viewing situations: a complex real scene viewed binocularly and a stereoscopic computer display of the same scene. The computer display is carefully constructed so all the standard depth cues—binocular disparity, texture gradients, occlusion, shading, etc.—are

Corresponding author: Martin S. Banks. martybanks@berkeley.edu. Address: 360 Minor Hall, UC Berkeley, Berkeley, CA 94720, USA.
Commercial relationships: none.

geometrically correct. Hence, the geometric patterns of stimulation striking the two eyes from the scene and display are the same. Psychophysical research (Backus, Banks, van Ee, & Crowell, 1999; Buckley & Frisby, 1993; Ellis, Smith, Grunwald, & McGreevy, 1993; Frisby, Buckley, & Duke, 1996; Frisby et al., 1995; Watt, Akeley, Ernst, & Banks, 2005) and experience with virtual-reality displays (Creem-Regehr, Willemsen, Gooch, & Thompson, 2005; Sahm, Creem-Regehr, Thompson, & Willemsen, 2005; Willemsen, Gooch, Thompson, & Creem-Regehr, 2007) suggest that the perceived 3D structure will differ: The depth in the computer display will generally appear flattened relative to the real scene from which it is derived (there are also cases in which the perceived depth will be exaggerated). In computer displays, focus cues—accommodation and blur in the retinal image-specify the depth of the display rather than the depicted scene. Those cues to flatness may affect depth percepts. Conventional 3D displays also create unnatural conflicts between vergence and accommodation. Those conflicts may affect the ability to fuse binocularly and may cause visual fatigue.

Here we describe work on the influence of accommodation and blur on binocular fusion, space perception, and visual fatigue. We also discuss the implications for vision research and the design and use of displays.

Focus cues and perceptual distortions

To understand 3D percepts, we should consider the array of cues the visual system uses to estimate 3D layout. There is considerable evidence that depth cues are combined in optimal or nearly optimal fashion to produce minimum-variance depth estimates (Hillis, Watt, Landy, & Banks, 2004; Jacobs, 1999; Knill & Saunders, 2003). Even supposedly ordinal depth cues are combined this way (Burge, Peterson, & Palmer, 2005). Assuming that the noises associated with cue measurement are independent and Gaussian distributed and all depths are equally likely, one can derive from Bayes' law a simple rule for producing the minimum-variance estimate (Cochran, 1937; Ghahramani, Wolpert, & Jordan, 1997):

$$\widehat{D} = \sum_i w_i D_i; \quad w_i = \sigma_i^{-2} / \sum_j \sigma_j^{-2}$$

(1)

where D_i *is* the relative depth (difference in distances between positions in the scene) specified by cue i, and σ_i is the standard deviation of that cue's estimate. Because the weights are proportional to the normalized inverse variances, more weight is assigned to less variable (more reliable) cues.

Stimuli in the real world contain numerous depth cues all specifying the same 3D layout. Computer displays present images on one surface: e.g., the phosphor grid for cathode-ray displays (CRTs). As a consequence, stimuli presented on computer displays contain some cues that specify the depth intended by the graphics engineer (we call these *simulated cues*) and some cues created by the display screen that specify the relative depth of the screen (namely, that it is flat) rather than the intended depth (*screen cues*). Screen cues include motion parallax due to the viewer's head movements relative to the screen and visible pixelization due to the discrete nature of the screen, but we concentrate here on one class of screen cues—focus cues —because they are likely to cause perceptual distortions (Buckley & Frisby, 1993; Frisby et al., 1995; Watt et al., 2005) and viewer fatigue (Ukai, 2007; Wann & Mon-Williams, 1997) and because they have proven difficult to eliminate (Akeley, Watt, Girshick, & Banks, 2004). Focus cues come in two forms.

1. Blur gradient in the retinal image. With real scenes, retinal blur varies consistently with changes in scene depth: The retinal image is sharpest for objects at the distance

to which the eye is focused and blurred for nearer and farther objects (Figure 1, lower left). The correlation between blur and depth in real scenes aids depth perception (Kuribayashi et al., 2006;Marshall, Burbeck, Ariely, Rolland, & Martin, 1996;Mather, 1997;Mather & Smith, 2000,2002;O'Shea, Govan, & Sekuler, 1997). In computer displays, focal distance is constant so all retinal-image points are sharp if the viewer's eye is focused at that distance and blurred if the eye is focused elsewhere. Consequently, the blur gradient produced by viewing a conventional 3D display specifies flatness (Figure 1, lower right).

2. Accommodation. As the eye looks around a real scene, neural commands are sent to the lens muscles to change focal power and thereby minimize blur for the fixated part of the scene. This is a depth cue (Mon-Williams & Tresilian, 2000; Wallach & Norris, 1963). As the eye looks around the simulated scene in a 3D display, the focal distance of the light does not vary, so the neural commands are constant and signal less depth (Figure 1).

If the estimates from individual cues are unbiased from the true values, the minimum-variance estimate from Equation 1 is

$$\widehat{D} = w_{\text{sim}} D_{\text{sim}} + w_{\text{foc}} D_{\text{foc}}, \tag{2}$$

where D refers to relative depth (the distance between one point and another), foc and sim refer to the relative depth specified by focus cues and by information other than focus cues (e.g., disparity, shading, perspective), and w_i are given by Equation 1. In the real world, relative depth specified by the two classes of cues would generally be the same, so the equation reduces to

$$\widehat{D} = D_{\text{sim}}. \tag{3}$$

A plot of perceived relative depth as a function of the depth specified by the available cues would yield a line of slope one. As a consequence, we generally see the depth in real scenes correctly, at least at close range (Mon-Williams, Tresilian, & Roberts, 2000); for counter-examples at long range, see Loomis, Da Silva, Fujita, and Fukusima (1992). With a computer-displayed stimulus, simulated cues indicate depth variation while focus cues indicate a relative depth of zero:

$$\widehat{D} = w_{\text{sim}} D_{\text{sim}}. \tag{4}$$

Thus, a plot of perceived relative depth as a function of the relative depth specified by cues other than focus cues yields a line of slope w_{sim}, which is less than 1 for $w_{\text{foc}} > 0$. Consequently, we usually experience depth compression in 3D displays (Buckley & Frisby, 1993; Watt et al., 2005; Willemsen et al., 2007).

Focus cues, fusion, and stereopsis

For a stimulus to be sharply focused on the retina, the eye must be accommodated to a distance close to the focal distance of the object. The acceptable range is the depth of focus, which is roughly ±0.3 diopters (D) under normal circumstances (Campbell, 1957; Charman & Whitefoot, 1977). Understandably, the blur caused by an accommodative error reduces the precision of stereopsis: With natural pupils, errors of 1D and 2D cause nearly two- and ten-

fold reductions in stereoacuity, respectively (Odom, Chao, & Leys, 1992; Westheimer & McKee, 1980; Wood, 1983). For a stimulus to be seen as single (i.e., fused) rather than double, the eyes must be converged to a distance close to the object distance. The tolerance range is Panum's fusion area, which is 15–30 arcmin (Ogle, 1932; Schor, Wood, & Ogawa, 1984). Thus, vergence errors larger than 15–30 arcmin cause a breakdown in binocular fusion and stereopsis is thereby disrupted (Julesz, 1971). Smaller vergence errors do not cause fusion to break down, but yield measurable reductions in stereoacuity (Blakemore, 1970). Therefore, fine stereopsis requires reasonably accurate accommodation and vergence. Figure 2A shows the range of acceptable focal distances (the depth of focus) and the range of acceptable vergence distances (Panum's area) when the viewer accommodates and converges to the same distance. The range of accommodation and vergence possible without excessive error in either is the *zone of clear single binocular vision* (Fry, 1939; Howard & Rogers, 2002; Morgan, 1944; green region in Figure 2B).

Accommodation and vergence responses are normally coupled. Specifically, accommodative changes evoke vergence changes (accommodative vergence), and vergence changes evoke accommodative changes (vergence accommodation) (Fincham & Walton, 1957; Martens & Ogle, 1959). In the real world, accommodation–vergence coupling is helpful because focal and vergence distances are almost always the same no matter where the viewer looks (Figure 1, left). One benefit of the coupling is increased speed of accommodation and vergence. Accommodation is faster with binocular viewing—where blur and disparity signals specify the same change in distance—than with monocular viewing where only blur provides a useful signal (Cumming & Judge, 1986; Krishnan, Shirachi, & Stark, 1977). Similarly, vergence is faster when disparity and blur signals specify the same change in distance than when only disparity specifies a change (Cumming & Judge, 1986; Semmlow & Wetzel, 1979). For these reasons, one expects that demanding stereoscopic tasks will require less time when the stimuli to accommodation and vergence are consistent with one another than when they are not.

In 3D displays, the normal correlation between focal and vergence distance is disrupted (Figure 1, right): Focal distance is now fixed at the display while vergence distance varies depending on the part of the simulated scene the viewer fixates. For reasons stated above, we expect that stereoacuity will be reduced in 3D displays compared with viewing situations that maintain the normal correlation between focal distance and vergence distance. We also expect that the time required to fuse a stimulus binocularly will be increased in conventional 3D displays.

Focus cues and visual fatigue

Prolonged use of conventional 3D displays produces viewer fatigue and discomfort (Emoto, Niida, & Okano, 2005; Takada, 2006; Takaki, 2003; Wann & Mon-Williams, 2002; Yano, Emoto, & Mitsuhashi, 2004). It has often been claimed that the symptoms are caused by the dissociation between vergence and accommodation that is required in such displays (Emoto et al., 2005; Ukai, 2007; Wann & Mon-Williams, 1997; Yano et al., 2004). Emoto et al. (2005) observed symptoms of fatigue when observers viewed stimuli with a larger conflict between the vergence and the focal distances. Wann and Mon-Williams (2002) reported a pre-test versus post-test change in the cross-link functions (AC/A and CA/C ratios) after prolonged viewing of a 3D display; they also observed an increase in fatigue. As we will make clear in the Discussion section (and foreshadow in the introduction to Experiment 4), factors other than the conflict in the vergence and the accommodative stimuli could have been responsible for the reported fatigue in all of the above-mentioned references. Therefore, to our knowledge, the link between the stimulus conflict and fatigue has never been directly tested.

The set of vergence and accommodative responses that can be achieved without discomfort is *Percival's zone of comfort,* which is about one-third the width of the zone of clear single binocular vision (Howard & Rogers, 2002; Morgan, 1944; Percival, 1920; yellow region in

Figure 2B). Stimuli in the real world (the circles in Figure 2B) fall within the comfort zone while many stimuli on 3D displays (the squares) do not. To fuse and focus the latter stimuli, the viewer must counteract the normal accommodation–vergence coupling, and the effort involved is believed to cause viewer fatigue and discomfort (Ukai, 2007; Wann & Mon-Williams, 1997). Accommodation and vergence cannot be set independently to arbitrary values because a change in vergence innervation affects accommodation and vice versa.

3D displays are used increasingly in several professions, particularly medicine, so it is important to determine if accommodative—vergence mismatches in fact cause the fatigue and discomfort. Determining whether or not such mismatches are the cause is a crucial first step toward figuring out how to minimize the problem.

Novel 3D display

Because of the problems associated with conventional 3D displays, there have been many attempts to construct displays that minimize the conflict between simulated and focus cues (and between vergence and accommodation) (Favalora et al., 2002; Lucente, 1997; McQuaide, 2002; Schowengerdt & Seibel, 2006; Sullivan, 2004; Suyama, Takada, Uehira, & Sakai, 2001; Takaki, 2003). These displays have not been widely used for a variety of reasons including the facts that they cannot be driven by conventional graphics hardware, and they cannot correctly render view-dependent lighting effects such as occlusions, highlights, and reflections. A central goal in the development of our display was to retain the ability to render view-dependent lighting effects and to use conventional graphics hardware while minimizing the conflict between simulated cues and focus cues. The display is shown in Figure 3 and described in the General methods section.

Preview

In the experiments reported here, we used the unique properties of our display to investigate how the relationship between vergence- and accommodation-specified distances affects visual performance and fatigue. Experiment 1 examined whether viewers can fuse a stereoscopic image more quickly when the conflict between simulated cues (in this case, disparity) and focus cues is minimized. Experiment 2 asked whether viewers achieve greater stereoacuity with brief stimulus presentations when the conflict is minimized. Experiment 3 examined distortions in perceived depth and how they are influenced by focus cues. Finally, Experiment 4 investigated the role of mismatches between the stimulus to vergence and the stimulus to accommodation in viewer fatigue and discomfort. We found in each case that minimizing the conflict between focus cues and disparity cues yielded a clear benefit.

General methods

The display is shown schematically in Figure 3A. In each eye's view, a mirror and two plate beam splitters are used to create a light field that is the sum of aligned images drawn at three image planes. We produce the binocular visual field using a periscope assembly. This creates a volumetric stereoscopic display because the light comes from sources at different distances. One way in which our display differs from other volumetric approaches is that we fix the eyes' positions. By knowing the observer's position, we can calculate each eye's view and display the correct disparities and preserve viewpoint-specific lighting effects like occlusion, specularity, and reflection. The display can present disparities and scene geometry at high resolution. It presents focus cues with relatively low resolution in depth because sensitivity to focus cues is far poorer than for spatial position (Campbell, 1957; Rolland, Krueger, & Goon, 1999). Figure 3A shows how we create three focal distances for each of the eyes by dividing the screen into six viewports.

Under typical viewing situations, depth of focus is ±0.25 to ±0.3 D (Campbell, 1957; Charman & Whitefoot, 1977), which corresponds to a range of 0.5–0.6 D around fixation. In our display, the image planes are placed slightly farther apart at 1.87 (far plane), 2.54 (mid plane), and 3.21 D (near plane), separations of 0.67 D (Figure 3B). The image–plane spacing is therefore just slightly greater than a standard observer's depth of focus. The monitor, an IBM T221 liquid-crystal display (LCD), has a maximum resolution of 3840 × 2400. At that resolution, pixels subtend 1.38, 1.09, and 0.80 arcmin at the near, mid, and far image planes, which for a standard observer is near the acuity limit. We drive the display with an Nvidia Quadro 4 900XGL graphics card. When the display was driven at the maximum resolution, the card could only support a 12-Hz refresh rate. Because of the sample-and-hold manner in which pixels are illuminated in LCDs, the images did not appear to flicker. For time-sensitive experiments, we ran the display at half resolution, 1920 × 1200, to boost the refresh rate to 41 Hz.

Alignment

Because the light rays approaching the eyes come from different distances, alignment of the eyes with the viewports is critical. We achieve this by first using a sighting device (Hillis & Banks, 2001) to adjust the bite bar (and therefore the observer's eyes) relative to the apparatus so that we can place the eyes in the appropriate position. Once the observer is in position, the separation between the viewing apertures is set equal to the observer's inter-ocular distance; we also set software parameters to the inter-ocular distance. We then fine-tune the alignment in software using a between-image–plane vernier alignment technique (Akeley et al., 2004). The alignment is accurate to within seconds of arc and ensures that light from the appropriate pixels on different image planes sums along lines of sight.

No need to track accommodation

With this display, we can simulate the effects of differential focus without tracking the observer's accommodation. To see this, consider a simple situation: We want to portray a small object at the distance of the far plane (53.6 cm) and another small object at the distance of the near plane (31.1 cm). For real objects positioned at those distances, accommodation to 53.6 cm would make the far object in focus and the near one out of focus. With accommodation to the near object, the reverse occurs. Exactly the same applies for our display because the light comes from different distances. The situation is more complicated for simulated object positions in between the planes. We discuss this in the next section.

To create the retinal images that best approximate images formed in real-world viewing, we render objects as unblurred on the image planes and let the observer's optics create the appropriate blur.

Depth-weighted blending

For all but the very unlikely case that the depth of a point in the scene coincides exactly with the depth of one of the image planes, a rule is required to assign image intensities to image planes. The simplest rule is the box filter (Figure 4, left): each point in the scene is drawn at the image plane to which it is closest. However, this approach produces blur discontinuities. For example, consider the situation in Figure 4 in which a line extends from a near image plane to a farther one. For a given accommodative state, the retinal-image blur of the line will have one value for parts that are drawn on one plane and another value for parts that are drawn on the second plane. This produces a visible discontinuity in the retinal image of the line. To minimize this problem, we use a tent filter (Figure 4, right). With this rule, the image intensity at each image plane is weighted according to the dioptric distance of the point from that plane, determined along a line of sight. This approach, which we call *depth-weighted blending,* eliminates the discontinuity. This blending technique is a significant technical development that can be applied to any multi-plane display.

Analysis of retinal-image formation

The retinal images formed by viewing real-world stimuli, stimuli on conventional 3D displays, and stimuli in our multi-plane display can differ substantially. In this next section, we explore how the retinal images differ across these three types of viewing situations to better understand the costs and benefits of different kinds of displays.

We measured the optics of an individual eye (the first author's left eye) using a Shack–Hartmann wavefront sensor. Optical aberrations were represented by Zernike polynomials from which we computed the *point-spread function* (PSF; the retinal image created by a point source) for objects at various distances with different accommodative states. We assumed that the aberrations do not change over small visual angles and therefore that we could calculate the retinal image from the convolution of a small object and the PSF. In order to calculate the PSF for varying amounts of defocus, we used the Zernike aberrations measured at one particular accommodative state and then added an appropriate amount of Zernike defocus. This assumed that the higher-order aberrations did not change with accommodation (Cheng et al., 2004).

Although we are ultimately interested in image formation with natural, broadband stimuli, it is instructive to study spatial sinusoids because broadband stimuli can be synthesized from sinusoids and because the primary effect of defocus on sinusoids is only a reduction in contrast. The upper, middle, and lower rows of Figure 5 show retinal-image contrast for sinusoids in the real world, on a conventional 3D display, and in our volumetric display. The left, middle, and right columns show the plots for stimulus spatial frequencies of 2, 6, and 18 cpd, respectively. The abscissas represent the real or the simulated focal distance of the stimulus in diopters. The ordinates represent the eye's focal distance in diopters. Colors represent the contrast ratio: retinal-image contrast divided by stimulus contrast (yellow representing the highest ratio and black the lowest).

Consider a real stimulus of 6 cpd (top center in Figure 5) at a distance of 2.5 D. As the observer accommodates from far to near (i.e., from 1.5 to 3.5 D, a vertical slice in the graph), the retinal-image contrast first increases to a maximum of 88.5% of the stimulus contrast and then decreases. As one would expect, focusing the eye at 2.5 D, the actual object distance, yields maximum retinal contrast. At a lower frequency of 2 cpd, the rise and fall of image contrast is shallower, and peak contrast is higher at 96.4%. At a higher frequency of 18 cpd, the rise and fall is steeper, and peak contrast is much lower at 48.2%. These plots represent the normal relationship between object distance, accommodative response, and retinal-image contrast.

Next consider a conventional 3D display (middle row in Figure 5). Because the distance to the display surface is fixed at 40 cm (2.5 D), the relationship between simulated object distance, accommodation, and retinal-image contrast is altogether different: retinal contrast is now maximized by accommodating to the distance of the display surface rather than to the object's simulated distance. Thus, to maintain a clear and single percept, the observer must hold accommodation fixed despite changes in simulated distance, and this requires the dissociation of accommodation and vergence.

Now consider our multi-plane display (Figure 5, bottom row). The three image planes are positioned at intervals of 0.67 D, so the workspace is a 1.33-D volume. When the simulated distance is at the distance of an image plane, the retinal contrast produced by viewing our display is identical to the contrast produced by viewing the real world. When the simulated distance is between planes, the retinal image is formed by a depth-weighted blend (Figure 4) of intensities from the two nearest planes. At 2 cpd, the blended image within the display's workspace is a nearly perfect approximation to the image produced by a real target. Importantly, retinal-image contrast is maximized by focusing at the simulated distance rather than at one of the image planes. At 6 cpd, the blended image is still a good approximation, and retinal contrast

is again maximized by focusing at the simulated distance rather than at one of the image planes. At 18 cpd, the blended image is a poorer approximation to the real world: the peak contrast occurs near the image planes rather than at the simulated distance. This analysis shows that the multi-plane approximation to the real world can be very good for spatial frequencies as high as 6 cpd (probably higher) with image planes separated by 0.67 D. The results depend, of course, on the eye's pupil size and aberrations: An eye with a smaller pupil and/or greater aberrations has greater depth of focus, so the volumetric display provides an even better approximation to the real world in such cases.

Figure 5 shows that our display creates retinal images that are an excellent approximation to the real world at low spatial frequencies, a good approximation at medium frequencies, and a poor approximation at high frequencies. An important question is how good is the approximation visually? In particular, how are blur perception and accommodation affected in multi-plane displays? Although the retinal blur created by our display is an approximation to the blur created by real stimuli, the ability to distinguish stimuli presented at different simulated distances may be similar in the two cases because perceived blur is affected most strongly by medium spatial frequencies (Granger & Cupery, 1972;Walsh & Charman, 1988) where the multi-plane approximation is good. Furthermore, we suspect that accommodation to multi-plane and real-world stimuli will be similar because human accommodation is controlled primarily by medium spatial frequencies (4–8 cpd; Mathews & Kruger, 1994;Owens, 1980;Phillips, 1974;Tucker, Charman, & Ward, 1986;Ward, 1987).

Experiment 1

When vergence and accommodative distances differ substantially, many viewers find it difficult to fuse a binocular stimulus. We conducted an experiment to determine how serious an effect this is for the vergence–accommodation conflicts that occur with conventional 3D displays. Specifically, we measured the effect of focus cues on the time needed to discriminate a cyclopean stimulus in a random-dot stereogram. We varied the conflict between vergence distance and focal distance, and for each combination of those distances, we found the shortest stimulus duration at which the observer could reliably determine the orientation of the cyclopean stimulus. Because of the coupling between accommodation and vergence, we expected that observers would perform better when vergence and focal distances were equal to one another.

We previously conducted an experiment similar to this (Akeley et al., 2004), but that experiment was limited in its ability to reveal the consequences of vergence–focal conflicts on binocular fusion. First, the test stimulus always required that the observer diverge the eyes, and there are two problems with that:

1. the observer could anticipate that the test stimulus would jump to a greater distance and start making the divergence movement before the stimulus actually appeared; and

2. we could not determine the consequences of vergence–focal conflicts when convergence was required instead of divergence.

Second, the test stimulus consisted of two stereoscopically defined planes, a red one and a green one. The task was to indicate whether the red or green plane was nearer. Such a task does not require accurate accommodation so it is not well designed to reveal the consequences of vergence–focal conflicts. Experiment 1 was designed to overcome both of the problems with the experiment reported by Akeley et al. (2004).

Methods

Observers—Three observers participated: ARG (29 years old), DS (23), and BGS (19). ARG was an author; DS and BGS were unaware of the experimental hypotheses. All had normal stereopsis as assessed by the Titmus Stereo Test. They wore their usual optical corrections. We did not test older observers because they are much more likely to have reduced accommodative range due to presbyopia.

Stimuli—The stimuli were random-dot stereograms depicting sinusoidal corrugations in depth. We wanted to make sure that vergence and accommodation would have to be reasonably accurate to perform the task, so the stereograms had a high dot density (45 dots/deg^2) and high corrugation frequency (1.35 cpd). With these values, relatively small errors in accommodation or vergence affect performance (Banks, Gepshtein, & Landy, 2004; Odom et al., 1992). The corrugations were oriented ±15 deg from horizontal, and observers indicated which of the two orientations was presented on each trial. The peak-to-trough disparity was 4.9 arcmin at all viewing distances. The stimulus was presented in a virtual circular aperture with a diameter of 4.2 deg; the surround was black. We verified that the orientation-discrimination task could not be performed monocularly. We operated the display at half resolution (1920 × 1200 pixels) to achieve a refresh rate of 41 Hz thereby allowing fine adjustment of the display time.

At the beginning of each trial, a fixation target was presented on the mid image plane (Figure 6A). The test stimulus was then presented at various vergence and focal distances. We presented two types of stimuli: Those in which the focal distances corresponded to image planes (non-blended stimuli) and those in which the focal distances were between image planes (blended stimuli). Testing with blended stimuli provides an important check that the use of depth-weighted blending does not disrupt the fusion and interpretation of stereo stimuli. Specifically, we wanted to determine whether such blending causes an unforeseen disruption of stereovision.

Non-blended test stimuli were presented with the vergence distance at the near, mid, or far plane and the focal distance at the near, mid, or far plane. Several combinations of vergence and focal distance were presented as shown in Figures 6B and 6C (conditions 1, 3, 4, 5, 6, 7, 9, A, C, D, and F). The conflicts in the stimuli to vergence and accommodation were 0, ±0.33, ±0.67, or ±1.33 D.

Blended test stimuli were presented with the vergence distance and focal distance at a position between the mid and near planes ("near–mid distance;" 34.8 cm) or between the mid and far planes ("mid–far distance;" 45.4 cm). We chose the dioptric mid-points between the image planes for the test of depth-weighted blending because the analysis in Figure 5 shows that our approximation is poorest at those distances. Those stimuli are depicted in Figure 6B (rightmost column) and Figure 6C (conditions B and E).

Procedure—The observer initiated a trial with a button press upon which the fixation stimulus (Figure 6A) appeared for 634–1415 ms, the duration being drawn from a random uniform distribution rounded to the nearest value given the refresh rate; the random interval discouraged the observer from making anticipatory eye movements. The fixation stimulus was a cues-consistent cross that helped the observer converge and focus at the mid image plane. The fixation stimulus was then extinguished, and the test stimulus appeared immediately at one of the various combinations of vergence and focal distance (in Figure 6, items 1, 3, 4, 5, 6, 7, 9, and A–F). At the end of the test stimulus presentation, a masking stimulus appeared to ensure that observers could not make judgments based on an after-image of the test stimulus. The mask was composed of dots randomly positioned in depth and was presented at the same vergence distance and focal distance as the test stimulus. Observers indicated the orientation of the corrugation with a key press. They were not provided feedback about the correctness of their response. The duration of the test stimulus was varied according to a 2-down/1-up adaptive

staircase. The staircase was run until 12 reversals had occurred. To minimize fatigue, observers were prompted to take a short break after every eight trials. Each experimental session contained all the vergence–focal distance combinations presented in a random fashion so that the observer could not anticipate which condition would be presented at a given time. A session lasted 293 trials on average. Observers completed two sessions, so they were tested twice on each vergence–focal distance combination.

Cumulative Gaussians were fit to the combined psychometric data for each vergence–focal combination using a maximum-likelihood criterion (Wichmann & Hill, 2001a, 2001b). We found the 75% correct point on each function and defined that as the time required to fuse the stimulus in each condition.

Results

The results for the non-blended stimuli are shown in Figure 7A, which plots the time the observers required to correctly identify stimulus orientation as a function of the difference between the vergence and the focal distances. Different observers exhibited different effect sizes, but in every case the required time decreased monotonically with decreases in the vergence–focal conflict. The minimum time of 400–500 ms occurred when the conflict was zero. Observers were somewhat faster at fusing the stimulus when the vergence jumped from the mid to the near plane than when it jumped from the mid to far plane.

These data show that differences in the vergence and the focal distances affect the ability to fuse stereoscopic stimuli. When vergence and focal distances change together, the stimulus can be fused and interpreted quickly. When vergence and focal distance do not change together, more time is required. These effects are surely caused by the cross-coupling between vergence and accommodation.

Figure 7B compares performance with the blended stimuli and the cues-consistent subset of the non-blended stimuli. It plots the duration required to correctly identify stimulus orientation as a function of the change in vergence and focal distance. The required time increased as the change in vergence and focal distance increased. We also measured the time to fusion for stimuli at locations A, C, D, and F in Figure 6. The fusion times for the 0.33 D conflicts did not differ from the cues-consistent blends; we omitted these data points for clarity. The results show that the blending technique used to simulate a cues-consistent stimulus does not hinder stereovision, at least in the difficult task of Experiment 1.

Experiment 2

We examined the consequences of inconsistency between focal and vergence stimuli using a second important criterion: the acuity of stereopsis. Specifically, we measured stereoacuity thresholds for briefly presented stimuli for different vergence–focal conflicts.

Methods

The observers were the same as in Experiment 1. The stimuli, procedure, and task were also the same with two exceptions. First, stimuli were always presented for 1 s rather than a variable duration. Second, the spatial frequency of the corrugations was varied in order to find the highest discriminable frequency. To enable the presentation of high corrugation frequencies, we set the display to its highest resolution (3840×2400 pixels), which required a reduction in refresh rate to 12 Hz.

As in Experiment 1, the experiment had two kinds of stimuli. With the non-blended stimuli, focal distances corresponded with image planes; the combinations of vergence and focal distances were the same as with the non-blended stimuli in Experiment 1 (1, 3, 4, 5, 6, 7, 9, A,

C, D, and F in Figure 6C). We presented these stimuli to determine if vergence–focal conflicts adversely affect stereo acuity. With the blended stimuli, the focal stimuli were depth-weighted blends between image planes; the combinations of vergence and focal distances were the same as with the blended stimuli of Experiment 1 (B and E in Figure 6C). We presented these stimuli to determine if depth-weighted blending adversely affects stereopsis.

Results

Figure 8A shows the results with the non-blended stimuli with near and far vergence distances. The highest corrugation frequency at which observers could reliably discriminate stimulus orientation is plotted as a function of the difference between the vergence and focal distances. The highest stereoacuity was obtained when the conflict was smallest. Thus, minimizing the vergence–focal conflict enables more precise stereopsis with relatively brief stimulus presentations. Stereoacuity was slightly higher for the near-vergence stimuli than for the far-vergence stimuli.

The differences in performance across the three observers are interesting. The effect of vergence–focal conflict was much larger in ARG than in DS and BGS. In Experiment 1, DS and BGS could fuse and interpret a 1.35-cpd corrugation with durations less than 1 s at all vergence–focal conflicts. ARG could not interpret the far-vergence stimuli unless the conflict was zero; she also exhibited a larger effect of conflict with the near-vergence stimuli. The stimulus duration in Experiment 2 was thus longer than DS and BGS required at 1.35 cpd, but not necessarily longer than ARG required. For these reasons, it is not surprising that ARG exhibited the largest effect of varying the vergence–focal conflict in Experiment 2. Presumably, we would have observed larger effects of vergence–focal conflict in DS and BGS if we had used shorter durations.

The results for the cues-consistent non-blended and blended stimuli are shown in Figure 8B. Stereoacuity is plotted as a function of the change in the vergence and focal distances. The data at ±0.33 D are from the blended stimuli presented between image planes, and the other data are from non-blended stimuli presented at planes. The figure shows only results with cues-consistent stimuli (vergence distance equals focal distance). We also measured stereoacuity for stimuli at locations A, C, D, and F in Figure 6 and found that performance with the 0.33 D conflicts did not differ from the cues-consistent blends; we omitted those data points for clarity. Stereoacuity worsened as the magnitude of the change in vergence and focal distance increased. Blending did not affect performance, suggesting that depth-weighted blending does not adversely affect fine stereopsis.

Experiment 3

We next examined how conflicts between the stimuli to vergence and accommodation affect depth perception: in particular, how such conflicts affect the perception of 3D shape. Horizontal disparity is an ambiguous determinant of 3D shape because a given pattern of horizontal disparity is consistent with an infinite set of shapes (Backus et al., 1999; Garding, Porrill, Mayhew, & Frisby, 1995; Rogers & Bradshaw, 1995). To uniquely determine shape, a distance estimate is also required. The process of using a distance estimate in the interpretation of the pattern of horizontal disparity is *disparity scaling*. The visual system uses both vertical-disparity and extra-retinal signals in disparity scaling; the weights assigned to those signals vary with viewing condition (Backus & Banks, 1999; Backus et al., 1999; Rogers & Bradshaw, 1995). The extra-retinal signals are provided by the sensed horizontal vergence of the eyes and also possibly by the sensed accommodative state. Conventional 3D displays present appropriate vergence signals, but inappropriate accommodative signals because the focal distance is fixed at the distance to the display. Consequently, the distance estimate used in disparity scaling might be erroneous when conventional displays are used.

Watt et al. (2005) found clear evidence for an influence of focal distance in disparity scaling. In their experiment, they varied focal distance by changing the physical distance between the observer and the display; they did so by moving the observer, so the observer could have been aware of the change in viewing distance by means other than blur and accommodative signals. In Experiment 3, we took advantage of the unique properties of our volumetric display to isolate accommodation and blur and then reexamined how focus cues *per se* affect disparity scaling and thereby affect depth perception. Thus, this experiment was designed to determine how much of the effects observed by Watt et al. (2005) were due to vergence–focal conflicts as opposed to other cues to distance.

Methods

Apparatus—The volumetric display was the one described in the General methods section. We will refer to the trials conducted with this display as the *volumetric condition*.

We also used a conventional 3D display as in Watt et al. (2005). This was a CRT mounted on a translating platform that allowed us to vary viewing distance precisely. We will refer to trials conducted with this display as the *conventional display condition*. The viewing distances with the conventional display were very similar to the focal distances with the volumetric display: 30 cm (near; 3.33 D), 40 cm (mid; 2.50 D), and 55 cm (far; 1.82 D). In this condition, the room was dimly lit so that the frame of the CRT was visible, and observers were well aware of changes in viewing distance. Dichoptic images were presented with liquid-crystal shutter glasses (Crystal Eyes, Stereographics, Inc.) synchronized to the CRT. We positioned observers with custom bite bars such that the midpoint of their inter-ocular axis was on the surface normal from the middle of the CRT. The display was spatially calibrated (Backus et al., 1999; Watt et al., 2005) so we could present disparities with an accuracy of better than 1/2 arcmin.

Observers—Eleven observers participated in preliminary testing. Three were not able to fuse the stimulus when the vergence–focal conflict was large, so we did not test them further. Two more could fuse all the stimuli but experienced significant eyestrain and withdrew from the study. The six remaining observers were DEJ (20 years old), BGS (22), DMH (23), AAA (23), JYL (23), and BV (31). Five were emmetropes. JYL was an uncorrected anisometrope with 0.25 D of myopia in the right eye and 1.5 D in the left eye. None wore optical corrections during the experiment. All had normal stereoacuity as assessed by the Titmus Stereo Test. One observer (DMH) was an author; the others were unaware of the experimental hypotheses. We did not test older observers because they are much more likely to have reduced accommodative range due to presbyopia.

Stimuli—As in Watt et al. (2005), the test stimulus was a random-dot stereogram of a vertical hinge in an open-book configuration (Figure 9). Observers indicated on each stimulus presentation whether the perceived hinge angle was more or less than 90 deg. Dot density was 3.3 dots/deg^2. The horizontal and vertical extents of the stimulus were determined by an elliptical clipping window with a height of 5 deg and an average width of 6 deg; a random component was added to the width. The texture gradients of the sides of the hinge were consistent with the disparity gradients, but the texture was not a reliable slant cue because of its low density. We verified this by presenting the stimuli monocularly to one observer and finding that he could not reliably perform the task. The background was black. The dots were produced by illuminating the red phosphor, which minimizes cross-talk from one eye's image to the other's.

We reduced the effectiveness of vertical-disparity signals so that observers would be sure to use extra-retinal signals for disparity scaling. We accomplished this by reducing the height of

the stimulus, which decreases the weight given to vertical disparity (Backus et al., 1999; Rogers & Bradshaw, 1995).

We randomly varied the slant of the angle bisector of the hinge by ±10 deg with a mean of 0 deg. Randomizing the base slant forced observers to base their judgments on the hinge angle rather than on the slant of one side of the hinge.

All of the stimuli in this experiment were presented on one image plane: i.e., depth-weighted blending was not used on the volumetric display. We did this to make the stimuli on the conventional display and volumetric display as similar as possible.

Procedure—Before every experimental session, observers completed a training session that was used to minimize bias in perceiving a right-angle hinge. During that session, the fixation cross and test stimulus were presented at the mid vergence and focal distance: 40 cm on the CRT and 39.4 cm on the volumetric display. On each training trial, the test stimulus was presented, and observers indicated whether the hinge angle appeared more acute or obtuse than a right angle. Auditory feedback about the correctness of the response was then provided. An adaptive, 1-down/1-up staircase procedure varied the hinge angle from trial to trial. Training sessions lasted about 3 min.

An experimental session began after a short break. Trials in those sessions started with the 1500-ms presentation of the fixation cross at the vergence and focal distance at which the hinge stimulus would appear. The hinge then appeared for 1500 ms (the fixation cross remaining visible). Then the cross and hinge were extinguished, and the observer indicated with a key press whether the hinge angle appeared more or less than 90 deg. A 1-down/1-up staircase adjusted the hinge angle over trials, terminating after 12 reversals. The initial step size of 24 deg was halved after each reversal until reaching a final step size of 3 deg. The resulting psychometric data were fit with a cumulative Gaussian using a maximum-likelihood criterion (Wichmann & Hill, 2001a, 2001b). The 50% point on the fitted function was the estimate of the hinge angle that appeared as a right angle. From those estimates, we calculated the *equivalent distance,* the distance at which the horizontal disparities in the stimulus specify a right angle. For a target straight ahead of the viewer, the equivalent distance is

$$D = \frac{\text{IPD}(\text{HSR}+1)}{2(\text{HSR}-1)}\tan(-\pi/4),$$

(5)

where IPD is the inter-ocular distance and HSR is the horizontal size ratio (a measure of relative horizontal disparity; Howard & Rogers, 2002).

There were two kinds of experimental sessions.

1. The *cues-inconsistent sessions were* conducted on both the volumetric and conventional displays. In those sessions, the focal distance was constant, and the vergence distance varied randomly from trial to trial. For some of those sessions, the focal distance was near, and the vergence distances were near, mid, and far. These conditions are numbered 1, 2, and 3 in Figure 6C. For some of the cues-inconsistent sessions, the focal distance was at the mid distance, and the vergence distances were at the near, mid, or far distance. These are numbered 4, 5, and 6 in Figure 6C. For the remainder of the cues-inconsistent sessions, the focal distance was at the far distance, and the vergence distances were at the near, mid, or far distance. These are numbered 7, 8, and 9 in Figure 6C. Note that 1/3 of the trials in the cues-inconsistent session

were actually no-conflict trials; that is, the vergence and focal distances were equal to one another 1/3 of the time.

2. The *cues-consistent session* was conducted on the volumetric display. In that session, the vergence and focal distances were always equal to one another. The distances were near, mid, and far (3, 5, and 7 in Figure 6C). By comparing observer behavior in the cues-consistent session with behavior for the three equivalent vergence–focal distance combinations in the cues-inconsistent sessions (the ones in which vergence distance equaled focal distance), we could determine the consequences of changing focal distance within a block of trials as opposed to between blocks of trials.

Each session consisted of six randomly interleaved staircases, two for each vergence distance. Every session was repeated at least three times, yielding at least 100 points per psychometric function.

Results

The results are shown in Figure 10, which plots equivalent distance as a function of vergence-specified distance. The left columns in each half of the figure show data from the conventional display and the right columns from the volumetric display. Different rows show data from different observers. The red, green, and blue points are the data from the near, mid, and far cues-inconsistent sessions, respectively; the color lines are regression fits to those data. The black points and lines in the right columns are the data from the cues-consistent session. If disparity scaling were based only on the vergence-specified distance and were done without error, the data would lie on the dashed diagonal lines. The slope of the data is always less than one, which reflects the well-known under-constancy in depth perception (Johnston, 1991). An effect of focal distance is evidenced by vertical separation of the data. There was an effect of focal distance in the conventional display data, which replicates the findings of Watt et al. (2005). There was also an effect of focal distance in the volumetric display data.

To assess the reliability of the effect of focal distance, we performed statistical analyses on the slopes of the regression lines in Figure 10. Recall that a slope of one indicates that the observer was taking the change in distance into account without error (i.e., had perfect depth constancy), and that a slope of zero indicates that the observer was not taking changes in distance into account at all (i.e., he/she was accepting the same pattern of horizontal disparities as specifying a right-angle hinge no matter what the distance was). The cues-inconsistent entries to the analysis were the average slopes of the three colored regression lines in Figure 10. We used two kinds of cues-consistent data for the analysis:

1. data from the volumetric display in which vergence and focal distance were equal to one another and changed form trial to trial; we call these *cues-consistent, within-session* data; and

2. a subset of the data from the cues-inconsistent sessions on the conventional and volumetric displays, the subset containing only those trials in which vergence and focal distance were equal to one another; we call these *cues-consistent, between-session* data.

The entries for cues-consistent, between-session were the slopes of the regression lines fit to the appropriate data subset. The entries for cues-consistent, within-session were the slopes of the black lines in Figure 10. On the volumetric display, the slopes were greater for the cues-consistent, between-session data than for the cues-inconsistent data; this result approached statistical significance (paired *t*-test, $p = 0.13$, one tailed). Also on the volumetric display, the slopes were significantly greater for the cues-consistent, within-session data than for the cues-inconsistent data (paired *t*-test, $p = 0.02$, one tailed). On the conventional display, the slopes were marginally significantly greater for the cues-consistent, between-session data than for the

cues-inconsistent data (paired *t*-test, $p = 0.09$, one tailed). Thus, observers exhibited more depth constancy on both displays when focal distance was consistent with vergence distance than when it was not. Said another way, presenting stimuli with appropriate focal distance increased the accuracy of depth perception.

The results are summarized in Table 1, which shows the slopes of the regression lines for each subject and condition. The cues-inconsistent values are the average slopes of the three line fits in Figure 10. The cues-consistent, between-session values are the slopes of the lines fitted through the cues-consistent trials in the different cues-inconsistent sessions. There were no statistically significant differences between the cues-inconsistent data collected with the conventional display and the volumetric display (paired *t*-test, $p = 0.50$). There were also no statistically significant differences between the cues-consistent, between-session data collected with the conventional display and the volumetric display (paired *t*-test, $p = 0.87$). Thus, the additional information available with the translating conventional display did not systematically affect perceived depth. This result indicates that focus cues *per se* were the key determinant of the improvement in depth constancy observed by Watt et al. (2005) when focal distance was the same as vergence distance.

There was significantly greater depth constancy in the volumetric display when vergence and focal distance were equal to one another and changed within a session than when they were equal to one another, but changed between sessions (paired *t*-test, $p = 0.05$). Thus, trial-to-trial changes in vergence and focal distance appear to improve the accuracy of disparity scaling.

In an attempt to understand why some observers responded differently to changes in focal and vergence distance than others did, we assessed vergence–accommodation coupling in every observer using standard clinical tests. AC/A (accommodative convergence over accommodation) is a measure of how much the eyes converge when the accommodative state of one eye is driven to different values. CA/C (convergence–accommodation over convergence) is a measure of how much the eyes accommodate when vergence is driven to different values. We found no correlation between these measurements and the perceptual results in Figure 10 and Table 1. Watt et al. (2005) made similar measurements with the same outcome.

It is possible that the between-subject variability was due to the use of a spectrally narrowband stimulus (red only). Narrowband stimuli are not an effective accommodative stimulus for many observers (Fincham, 1951), so we might have observed less between-subject variability if we had used a white stimulus.

Experiment 4

As we said earlier, conventional 3D displays are believed to cause fatigue and discomfort. Most researchers and engineers have assumed that the symptoms are caused by differences between the stimuli to vergence and accommodation because such differences require the viewer to uncouple vergence and accommodation (Emoto et al., 2005; Fry & Kent, 1944; Häkkinen, Pölönen, Takatalo, & Nyman, 2006; Howarth & Costello, 1997; Menozzi, 2000; Miles, Judge, & Optican, 1987; Wann & Mon-Williams, 2002; Yano et al., 2004). The evidence offered in support of this hypothesis is that viewers report more fatigue and discomfort when viewing 3D displays than when viewing 2D displays (Emoto et al., 2005; Häkkinen et al., 2006; Jin, Zhang, Wang, & Plocher, 2007; Yamazaki, Kamijo, & Fukuzumi, 1990; Yano, Ide, Mitsuhashi, & Thwaites, 2002). This observation, however, does not prove that vergence–accommodation conflicts cause fatigue and discomfort because there are several other important differences between viewing 2D and 3D displays; these include the eye wear required with 3D displays to separate the two eyes' images, the ghosting or cross-talk from one eye's image to the other's (Kooi & Toet, 2004), and the perceptual distortions that occur with 3D displays (Bereby-Meyer,

Leiser, & Meyer, 1999) and not with 2D displays (Vishwanath, Girshick, & Banks, 2005). To our knowledge, no one has shown that vergence–accommodation decoupling *per se* causes fatigue and discomfort. Two papers—Emoto et al. (2005) and Yano et al. (2004)—came closest. We review those papers in the Discussion section ("Focus cues and visual fatigue") and make clear why they were unable to show that conflicts in the vergence and accommodative stimuli cause visual fatigue.

Our display provides a unique opportunity to test the conflict hypothesis because with this display we can independently vary vergence and focal distances without any other changes in the stimulus or task. Experiment 4 tests the hypothesis directly.

Methods

Observers—Eleven observers, aged 23 to 31 years, participated. We measured their stereo vision using the Titmus Stereo Test and all were normal. They all wore their usual optical corrections during the experiment. All were unaware of the experimental hypotheses but were warned that they might experience fatigue and discomfort. We did not include older observers because they were likely to have decreased accommodative ability due to presbyopia.

Apparatus and stimuli—The experiment was conducted on the volumetric display in Figure 3. The test stimuli were disparity-defined sinusoidal corrugations as in Experiments 1 and 2. They were presented within a circular aperture with a diameter of 6.5 deg. Peak-to-trough disparity was 15 arcmin (three times greater than in Experiments 1 and 2) and spatial frequency ranged from 1 to 3 cpd with dot density roughly proportional to the square of spatial frequency. Observers could not perceive the corrugation waveform if they accommodated or converged inaccurately.

Procedure—We used a three-interval, forced-choice oddity task. A trial began with the presentation of a 600-ms fixation cross on the mid image plane (vergence distance = focal distance = 39.4 cm). Then three 1500-ms presentations of the test stimulus occurred separated by inter-stimulus intervals of 250 ms. Each presentation was marked with a brief audible beep. The spatial frequency of the corrugation was the same in the three intervals, but the orientation (±15 deg from horizontal) was different in one. The task was to identify the odd interval. We provided auditory feedback to indicate the correctness of the response. The next trial began immediately after the observer responded or 4.5 s after the end of the previous trial, whichever came first. Average trial duration was 6 s. No breaks were allowed once the experimental session had started. We used this task because good performance required that the observer accommodate and converge reasonably accurately in each of the three stimulus intervals, and this was very demanding visually.

Experimental sessions had 1230 stimuli presentations and lasted about 45 min; each observer went through two sessions on consecutive days. During the *cues-consistent* session, the focal and vergence distances of the test stimulus were equal to one another at the near, mid, and far image planes (3, 5, and 7 in Figure 6C); those distances were randomly selected for each of the three stimulus intervals in a given trial. During the *cues-inconsistent* session, focal distance was fixed at the mid plane throughout, and vergence distance was randomly assigned in each interval to the near, mid, or far plane (4, 5, and 6 in Figure 6C). Thus, 2/3 of the intervals in the cues-inconsistent session contained vergence–focal conflicts of ±0.67 D, whereas none of the intervals in the cues-consistent session contained conflicts. The cues-consistent and cues-inconsistent sessions were otherwise identical: same stimulus waveforms, same task, same timing, etc. Therefore, the only difference between the two types of sessions was the relationship between vergence and focal distances. By isolating the vergence–focal conflict, we were able to directly test whether differences between vergence and focal distances

contribute to fatigue and discomfort. The order of sessions for each observer was assigned randomly. The observer and experimenter were unaware of which session was being run on a given day, so the experimental design was double blind.

At the beginning of each session, the experimenter read a script explaining that we were examining visual fatigue associated with viewing 3D displays. The script also described the procedure without revealing how the experimental sessions differed. In keeping with the human subjects protocol, the experimenter informed subjects of their right to withdraw from the study at any time. A session began with five minutes of training in the task. Training trials were the same as experimental trials except there was no conflict in the training trials (vergence and focal distances were both at the mid image plane). After training, observers were given a short break and then an experimental session began.

We wanted to assess the subjective experience of fatigue and discomfort. To do this, we had observers complete two questionnaires. They completed a symptom questionnaire at the end of each session (Figure 11, top). The questionnaire was modeled after one developed by Sheedy and Bergstrom (2002). It had five questions:

1. How tired are your eyes?

2. How clear is your vision?

3. How tired and sore are your neck and back?

4. How do your eyes feel?

5. How does your head feel?

In each case, observers indicated the severity of their symptoms at that moment. Questions 1, 2, 4, and 5 concerned symptoms that are believed to be affected by the vergence–focal conflict in 3D displays. Question 3 concerned the neck and the back, which should be unaffected by the conflict. We added that question to check whether participants were responding specifically to the queried symptom in each question or more generally.

Observers also completed a display-evaluation questionnaire after the second session (Figure 11, bottom). It had four questions.

1. Which session was most fatiguing?

2. Which session irritated your eye the most?

3. If you felt headache, which session was worse?

4. Which session did you prefer?

After observers completed the two sessions and the questionnaires, we invited them to return for another two sessions. If they agreed, the cues-consistent and cues-inconsistent sessions were presented again on consecutive days but in the opposite order from what they encountered before. Of the 11 observers, 9 returned for a second set of sessions with the ordering reversed. Three of these subjects were participants in a pilot experiment, and only their second set of data is reported. Two observers declined to return for the second set of sessions. Our results are thus drawn from a total of 17 sets of questionnaires.

We analyzed the percent-correct data for the orientation-discrimination task. Performance was nearly always at 80–100% correct, which means that observers were generally attentive and able to do the task. There was no systematic decline in performance over time.

Results

All subjects reported fatigue and/or discomfort at the end of every session, so the experiment was effective in producing the symptoms of interest. Figure 12 shows the average reported symptoms on the symptom questionnaire: The more severe the reported symptom, the higher the plotted bar. Orange and blue bars represent the reported symptoms for the cues-consistent and cues-inconsistent sessions, respectively. For questions 1, 2, 4, and 5, the cues-inconsistent symptoms were significantly more severe than the cues-consistent symptoms (Wilcoxon signed-rank test, $p < 0.025$, one tailed). The severity of reported symptoms did not differ significantly for question 3. Thus, when vergence and focal distance were not always the same, subjects experienced more symptoms associated with the eyes and the head; as expected, they did not experience more symptoms associated with the neck and back.

Figure 13 shows the results for the display-evaluation questionnaire in which subjects were asked to compare their symptoms in cues-consistent and cues-inconsistent sessions. Higher values indicate more favorable ratings for the cues-consistent session in which vergence and focal distance maintained equal values even as they changed. For all four questions, the cues-consistent session received more favorable ratings: The difference was statistically significant for questions 2, 3, and 4 (Wilcoxon signed-rank test, $p < 0.025$, one tailed) and question 1 ($p < 0.05$, one tailed). Subjects experienced the two sessions in random order, so the observed difference reflects preferences for the type of session and not the order in which they were presented. The results from the display-evaluation questionnaire make clear that subjects had more favorable experiences when vergence and focal distance changed together than when vergence changed and focal distance was fixed as in conventional 2D displays.

The fatigue and discomfort results in Figures 12 and 13 may actually under-represent the size of the difference between the two experimental sessions. The two subjects who reported the most severe symptoms during the cues-inconsistent session declined to return for the voluntary second pair of sessions. Thus, unlike the other subjects, they only contributed data from two sessions as opposed to four.

These results are, we believe, the first demonstration that mismatches in the stimuli to vergence and accommodation cause visual fatigue and discomfort. As such, our findings are important for a variety of applications of 3D technology. We will return to this topic in the Discussion section.

Discussion

Vergence–focal compatibility and visual performance

In Experiment 1, we found that the time required to discern the cyclopean stimulus in a random-dot stereogram was minimized when the vergence and focal distances were equal to one another. In Experiment 2, we found that finer depth corrugations could be discriminated when vergence and focal distances were the same. Both of these effects are almost certainly byproducts of the cross coupling between vergence and accommodation (Fincham & Walton, 1957). When the demand in the vergence stimulus is equal to the demand in the accommodative stimulus, the cross links facilitate rapid and accurate responses, so accommodation and vergence attain their end states more rapidly than when the demands are unequal (Cumming & Judge, 1986; Krishnan et al., 1977; Semmlow & Wetzel, 1979).

Accommodative responses to changes in object distance consist of two components: a slow component driven by blur in the retinal image and a fast component driven by binocular disparity, the stimulus to vergence (Khosroyani & Hung, 2002). In natural viewing, the changes in vergence and focal distances are the same (meaning that the vergence and accommodative demands are equal), and the slow and fast components work together to produce a relatively

rapid and accurate accommodative response. When viewing conventional 3D displays, the vergence and accommodative demands generally differ, so the slow and fast components attempt to drive accommodation to different values. The vergence-driven fast component produces a rapid response to an accommodative state that does not minimize blur. The blur-driven slow component senses the error and feeds the cross-coupled system to correct the overshoot or undershoot produced by the fast component. As a consequence, the response is slower than when the vergence and focal stimuli are consistent; indeed, when the vergence and focal stimuli differ, accommodation occasionally oscillates (Cumming & Judge, 1986; Semmlow & Wetzel, 1979). These effects have been observed in viewers of conventional 3D displays (Torii, Okada, Ukai, Wolffsohn, & Gilmartin, 2008).

The vergence response is also affected by the consistency of the vergence and focal stimuli, but the effects are smaller than with the accommodative response (Cumming & Judge, 1986; Torii et al., 2008). In a study of 3D displays, Emoto et al. (2005) presented two kinds of stimuli: ones in which the vergence and focal distances were consistent with one another and ones in which vergence distance was varied and focal distance was not. Observers reported many more fusion failures (i.e., diplopia) with the latter stimuli, which indicates that vergence was less accurate when vergence–focal conflicts were present.

The stimulus in Experiments 1 and 2 required accurate accommodation because retinal-image blur impedes the ability to discern high-frequency depth corrugations (Banks et al., 2004; Odom et al., 1992). Thus, conflicts between the stimulus to vergence and the stimulus to accommodation undoubtedly affected the speed and the accuracy of the accommodative response thereby producing some, if not all, of the effects we observed.

It is interesting to note that the accommodation response produced by the vergence–accommodation system is quite dependent on the spatial-frequency content of the stimulus. When the stimulus contains high frequencies, changes in retinal-image blur are readily detected, so the blur-driven slow component drives the cross links yielding corrections, even oscillations, for hundreds of milliseconds (Okada et al., 2005). In contrast, when the stimulus contains only low spatial frequencies, changes in retinal-image blur are hard or impossible to detect, so the blur-driven component does not drive the system to correct accommodative errors; rather, accommodation settles relatively quickly near the value specified by the vergence stimulus (Okada et al., 2005). Because of this, the need to minimize the vergence–focal conflict becomes greater and greater as the spatial resolution of the display is increased.

Many viewers cannot fuse a binocular stimulus with a vergence–focal conflict. We documented this in the preliminary testing for Experiment 3. We presented our stimuli to 11 young subjects. Three could not fuse most of the stimuli. Another two could fuse the stimuli but complained that doing so was too fatiguing. Only six subjects could fuse all the stimuli without significant fatigue and discomfort, and they were chosen for the main experiment. All of the disqualified subjects had normal binocular vision and could also fuse binocular stimuli in the natural environment across a wide range of distances. They only had problems with the experimental stimuli in which vergence–focal conflicts were present. The fact that nearly half of observers, all of whom were young adults with normal binocular vision, could not readily fuse stimuli with vergence–focal conflicts is a reminder that researchers and 3D display engineers should minimize the conflict in their displays to increase the number of people who can participate.

Focus cues and depth perception

In Experiment 3, we investigated the consequences of vergence–focal conflicts on the perception of 3D shape. Although the effect was small, perceived depth was consistently more accurate when the vergence and the focal distances were the same as opposed to when they

differed. This effect was presumably caused by changes in the distance estimate the visual system uses to scale horizontal disparities (Garding et al., 1995; Watt et al., 2005).

The fact that focus cues can affect 3D percepts has significant implications for psychophysical research on depth perception. The great majority of experiments are conducted with conventional 3D displays in which focus cues specify the flat surface of the display rather than the simulated depth of the experimental stimulus. Consider, for example, research on the perception of structure from motion. The retinal-image motion caused by relative movement between the viewer and the object creates a compelling 3D impression, but in laboratory experiments, the judged depth is often smaller than the simulated depth (Braunstein, Liter, & Tittle, 1993; Caudek & Proffitt, 1993; Domini & Caudek, 1999; Loomis & Eby, 1989; Todd & Bressan, 1990). For instance, Hogervorst (1998) presented monocular structure-from-motion stimuli that simulated hinges similar to the ones in our Experiment 3. As in our experiment, observers judged the angle between the sides of the hinge. They consistently overestimated the angle (underestimated the depth). The results were consistent with a Bayesian model incorporating optic flow measurements and their associated noises. However, depth underestimation would also be expected if focus cues affected observers' percepts because such cues specify the flat surface of the display rather than the varying distance of the sides of the simulated hinge.

Hogervorst and Eagle (2000) tested for effects of focus cues by redoing part of the experiment with a pinhole pupil. They reasoned that a pinhole would render focus cues uninformative, so if focus cues had contributed to the depth underestimation in their main experiment, they would observe less underestimation with a pinhole than with a natural pupil. However, pinholes do not necessarily render focus cues uninformative; instead they may cause the retinal blur and accommodative signals to be interpreted as specifying flatness (Frisby et al., 1996; Watt et al., 2005). For this reason, using pinholes with conventional displays is not an adequate test for the influence of focus cues. By not including signals that may well have affected observers' percepts, Hogervorst and Eagle may have misinterpreted their data yielding an erroneous theoretical account.

Todd and colleagues have repeatedly observed non-veridical depth estimation when 3D structure is specified by disparity, motion, pictorial information, and combinations of those cues (reviewed by Todd, 2004). Their stimuli are also generally presented on computer displays, so focus cues may have contributed to the compression in perceived depth. Todd and colleagues have discounted this concern by noting that non-veridicality is frequently observed with real stimuli as well (e.g., Baird & Biersdorf, 1967). But to gain a full understanding of depth perception, we need to quantify *what* the depth percepts are, not just whether or not they are veridical. Quantification could be improved by evaluating the contribution of focus cues in such experiments.

Focus cues and visual fatigue

As we noted in the Introduction section, many authors have stated that conflicts in the stimuli to vergence and accommodation are the cause of the well-known visual fatigue and discomfort associated with conventional 3D displays (Howarth & Costello, 1997; Wann & Mon-Williams, 2002). Despite this widespread belief, no one has proven (with the appropriate control conditions) that the stimulus conflict causes fatigue and discomfort. In our opinion, the two reports that come closest are (Emoto et al., 2005; Yano et al., 2004).

In a study involving the manipulation of vergence and focal demands while viewing displayed images, Emoto et al. (2005) had subjects view a 3D display under different viewing conditions. There was a range of disparities in the scene. The absolute disparity of the entire scene was changed with variable-power prisms. In one condition, the absolute vergence stimulus was

varied periodically while the focal stimulus was fixed. In a second condition, the vergence and the focal stimuli were periodically changed in a fashion that made them somewhat similar to one another. Subjects reported the most fatigue when the prism power was changing, but the differences across conditions were not statistically significant. Emoto et al. (2005) concluded that vergence–accommodation conflicts cause visual fatigue. As we implied earlier, this conclusion is not warranted for three reasons.

1. The 3D movie that served as their stimulus presented multiple potential vergence distances, so Emoto and colleagues could not know the size of the conflict between the stimulus to vergence and the stimulus to accommodation for any of their experimental conditions: It would have depended on where the subject was looking in the virtual scene.

2. The task—counting the instances of a particular Japanese character—could have been performed monocularly and therefore did not require that the observer resolve the vergence–accommodation conflict.

3. The vergence stimulus was varied by changing the power of prisms in front of the observers' eyes and such prisms produce spatial distortions that can cause fatigue (Kooi & Toet, 2004).

Yano et al. (2004) presented stereo images of a page of text on a conventional 3D display. The subjects were instructed to read text. The focal distance of the text was constant, but the disparity-specified distance of the text differed from one experimental session to another. At the end of each session, they assessed subjects' fatigue and discomfort. The severity of the reported symptoms grew monotonically with the magnitude of the conflict between the focal and the disparity-defined distances of the text. Yano and colleagues concluded that vergence–accommodation conflicts cause visual fatigue and discomfort in conventional 3D displays. This conclusion is not justified for two reasons. First and most important, the vergence–focal conflict co-varied with the vergence demand, so one cannot determine from the results whether the conflict or the vergence demand caused the fatigue. Second, the subjects' task did not require stereopsis; indeed, it could have been performed monocularly.

Thus, to our knowledge, it has never been shown in an experiment with appropriate control conditions that differences in the stimuli to vergence and accommodation actually cause viewer fatigue and discomfort. We performed a direct test of the conflict hypothesis in Experiment 4. We isolated the variable of interest—the vergence–focal conflict—while keeping all other aspects of the stimulus and the procedure the same. Furthermore, the task—discriminating the orientation of a corrugation in a random-element stereogram—could only be performed binocularly and thus required stereopsis. We believe this is the first demonstration that this conflict is a cause of visual fatigue and discomfort.

The fatigue and discomfort produced by vergence–focal conflicts are presumably byproducts of the cross coupling between vergence and accommodation (Fincham & Walton, 1957). When disparity is present, commands are sent to the extra-ocular muscles to modify vergence and to minimize the disparity. When blur is present, commands are sent to the ciliary muscles to adjust accommodation and minimize the blur. However, vergence and accommodation are additionally complicated because of the vergence–accommodation cross coupling: changes in disparity affect accommodation via vergence commands and changes in blur affect vergence via accommodative commands. When the vergence and the focal stimuli are consistent with one another, the commands are compatible and the two systems can adjust to appropriate states relatively quickly. When the vergence and the focal stimuli differ, the commands are no longer compatible. Presumably, the difficulty in resolving the incompatibility is an important contributor to fatigue and discomfort. Consistent with this argument, Wann and Mon-Williams

(2002) observed a change in the coupling between vergence and accommodation (i.e., changes in the AC/A and CA/C ratios).

Transient versus sustained changes in focal distance

In Experiment 3, there were two kinds of experimental sessions presented with the volumetric display:

1. sessions in which the focal distance was the same for the whole block of trials and vergence distance changed from trial to trial (in 1/3 of those trials, the focal and the vergence distances were the same), and

2. sessions in which the focal distance and vergence distance were always equal to one another and changed from trial to trial.

We compared performance in the cues-consistent trials from those two sessions (Table 1, cues-consistent, between-session trials and cues-consistent, within-session trials). When we did so, we found that shape perception was more accurate in the cues-consistent, within-session trials than in the cues-consistent, between-session trials. This seems surprising because with trial-to-trial changes in focal distance, the system has less time to accommodate to the new distance and to estimate accommodative state than it has when focal distance is fixed within a block. We offer two hypotheses to explain this observation.

First, the cues-consistent trials in the first session were presented among cues-inconsistent trials in which the vergence and the focal distances differed. Perhaps the presence of those inconsistent trials caused changes in the vergence–accommodation system (i.e., adaptation of the cross links) that affected performance in the cues-consistent trials (Schor & Tsuetaki, 1987).

Second, perhaps the visual system can take transient changes in accommodative state into account better than sustained changes. There are examples of similar phenomena in the literature. For example, consider the problem of judging the perceived direction of a target relative to the head when the eyes are in eccentric gaze. To estimate direction correctly, the visual system must measure the position of the target on the retina and the position of the eyes relative to the head. Direction judgments are more accurate shortly after the eye movement is made than after the eyes are held in the eccentric position for a while (Paap & Ebenholtz, 1976). Thus, transient changes in eye position are taken into account better than sustained changes, and the same may apply to accommodation.

Retinal images formed by volumetric displays and the real world

The retinal images formed when viewing our volumetric display compared to the images formed when viewing the real world depend on several factors. Some of those factors are properties of the display: the form of the filter used for determining image intensities (Figure 4) and the separation between and number of image planes. Some factors are properties of the viewer: pupil diameter, accommodative accuracy, and the optical quality of the eye. Finally, some factors are stimulus properties: the spatial-frequency content and contrast of the object. It is important to know how the abovementioned factors affect the quality of our simulation of real-world viewing. Here we examine the effects of viewer and stimulus properties. To do so, we performed an analysis like the one in Figure 5. Specifically, we determined how retinal-image contrast varies with accommodative response for real objects and for virtual objects presented in our volumetric display.

Figure 14 plots contrast ratio (retinal-image contrast divided by object contrast) as a function of accommodative response to a stimulus rendered at the dioptric mid-point between two image planes in the volumetric display and for a real-world stimulus presented at the same focal

distance. We chose the dioptric mid-point between image planes because our approximation to the real-world is poorest at that position: If the approximation is good there, it will be at least as good at other positions between the image planes. In each panel, the dashed lines represent the contrast ratios for a real-world stimulus, and the solid lines represent the ratios for the same stimulus presented with depth-weighted blending in the volumetric display; the shading represents the difference in the ratios. As before, we represented the eye's optical aberrations with Zernike polynomials from which we computed the point-spread function (PSF). We again assumed that the higher-order aberrations measured at one accommodative state do not change with accommodation (Cheng et al., 2004).

The optical quality of observers with nominally normal vision can vary substantially (Thibos, Hong, Bradley, & Cheng, 2002). Depth of focus is greater in highly aberrated eyes than in eyes with high-quality optics (Chen, Kruger, Hofer, Singer, & Williams, 2006), so our approximation should be better when the observer has an aberrated eye. Figure 14A illustrates this by showing how the eye's aberrations affect contrast ratios in our display and the real world. Pupil diameter is 4 mm, and stimulus spatial frequency is 6 cpd. The red curves are calculations for a diffraction-limited eye (i.e., an eye with no aberrations). The green and blue curves are from observers with typical and greater-than-typical aberrations, respectively. As expected, the approximation to the real world is best for the highly aberrated eye and poorest for the diffraction-limited eye.

The depth of focus of the visual system is roughly inversely proportional to pupil diameter (Charman & Whitefoot, 1977; Green, Powers, & Banks, 1980), so our approximation to the real world will improve as pupil size decreases. Figure 14B demonstrates this by showing how the observer's pupil diameter affects contrast ratios in our display and the real world. The observer has typical optical aberrations relative to a population of young observers with normal vision (Thibos et al., 2002). The stimulus is a 6-cpd sinusoidal grating. We concentrate on that spatial frequency because, as we said earlier, blur perception and accommodation are determined primarily by spatial frequencies near that value (Owens, 1980; Walsh & Charman, 1988). The red, green, and blue curves represent contrast ratios for pupil diameters of 2, 4, and 6 mm, respectively. As expected, the volumetric display approximates the real world very well at 2 mm, reasonably well at 4 mm, and not as well at 6 mm.

Depth of focus is roughly inversely proportional to the spatial frequency of the stimulus (Green et al., 1980), so our approximation should be better at low than at high spatial frequencies. Figure 14C confirms this expectation by showing how stimulus spatial frequency affects contrast ratios in our display and the real world. Pupil diameter is 4 mm, and the eye has typical aberrations. The red, green, and blue curves represent contrast ratios for 2, 6, and 18 cpd, respectively. As expected, the approximation to the real world is excellent at 2 cpd, reasonable at 6 cpd, and poor at 18 cpd.

These calculations show that our volumetric display produces the best approximation to the real world when the depth of focus of the visual system is relatively large (small pupil, low spatial frequency, aberrated eye), and that it produces the poorest approximation when depth of focus is small (large pupil, high frequency, unaberrated eye). Experimenters and applications engineers should consider these issues when utilizing displays in basic vision research and in applications.

Practical guidelines

Basic researchers and engineers have to make a variety of decisions when they set up their experiments or applications. In most cases, they wish to make their visual displays as useful and non-fatiguing as possible, and they generally want their stimuli to produce percepts that

are faithful to the real world. The work presented here allows us to provide some guidelines for setting up and designing displays and for creating visual stimuli.

First, how does one minimize perceptual distortions due to focus cues? At this point, we do not have sufficient understanding of how depth percepts are affected by retinal blur and accommodation and by vergence–accommodation conflicts to state firm rules. There are, however, some general guidelines.

1. Use long viewing distances when possible because focus cues have less and less influence as the distance to the display increases; there should be little influence beyond 1 m (Watt et al., 2005).

2. Minimize the conflict between the vergence and the focal distances because conflicts cause inappropriate disparity scaling and failures of binocular fusion (Watt et al., 2005; Experiment 3 in the current paper). One can minimize the vergence–focal conflict by matching the simulated distance in the display and focal distance as well as possible.

3. Maximize the reliability of other depth cues in the stimulus because the influence of focus cues should decrease as its relative reliability decreases (Equation 1).

Second, how does one minimize visual fatigue and discomfort when using conventional 3D displays? We do not understand all of the determinants of fatigue and discomfort, but again there are some reasonable guidelines.

1. Minimize the vergence–focal conflict by setting the physical distance to the display as close as possible to the disparity-specified distance of the important parts of the virtual scene (Experiment 4 in the current paper). If the physical distance cannot be changed, ophthalmic lenses can be used to make the focal distance compatible with the vergence distance. Of course, one cannot in general set the vergence–focal difference to zero in conventional displays because the virtual scene will contain a variety of disparities, so the vergence stimulus will vary as the viewer fixates different parts of the scene.

2. When reasonable, minimize the consequences of vergence–focal conflicts by making an existing conflict less salient. This can be accomplished by increasing the distance to the display and to the virtual scene. It can also be accomplished by attenuating high-spatial frequency content from the images (Okada et al., 2005), but this is generally undesirable for other reasons.

The importance of focus cues, including vergence–focal conflicts, undoubtedly varies with the viewer's age. With increasing age, the range of accommodation decreases. By age 40–50 years, most viewers have become presbyopic. When presented with a volumetric display with near-correct focus cues, presbyopes will be unable to accommodate differentially to stimuli in different image planes. They will therefore experience appropriate depth-dependent blur but will be unable to reduce blur by accommodating differentially. In our experience, presbyopes do not benefit much, if at all, from the presentation of near-correct focus cues as in our display.

We close by demonstrating another effect of blur in the interpretation of images. When we look at an object in the natural environment, vergence and accommodation adjust to the appropriate distance. Therefore, the retinal images of the fixated object have no disparity and are well focused. As the viewer continues to look at the near object, background objects create retinal images with disparity and blur. In many simulations of that viewing situation, the disparity of the background is recreated correctly, but blur is not because the background objects are drawn as unblurred on the display.

Figure 15 provides examples of those two viewing situations. The stereograms depict two frontoparallel planes of sticks. In the upper stereogram, the sticks are all drawn unblurred. When you fixate the foreground (aided by the fixation marker), many of the background sticks appear double (or create false matches that produce various apparent stick orientations). The lower stereogram is identical to the upper except that the background sticks are blurred to simulate the effect of focusing on the foreground sticks. As you fixate the foreground, the diplopia associated with the background sticks is less apparent and the number of false matches is reduced. Thus, the blur induced by non-fixated background objects may reduce the salience of the diplopia that occurs with sharply rendered 3D simulations. This illustrates another potentially useful function of presenting correct or near-correct focus cues in 3D displays.

Conclusions

Conventional 3D displays present images on one surface, so focus cues—accommodation and blur in the retinal image—specify the depth of the display rather than the depths in the simulated scene. In addition, the vergence stimulus in conventional 3D displays varies depending on where the viewer looks in the simulated scene, but the focal distance remains fixed; the difference in those distances requires the viewer to uncouple vergence and accommodation. We examined how visual performance, perceived 3D shape, and visual fatigue are affected by the vergence–focal conflicts that arise in conventional 3D displays. To do so, we used a novel 3D display that presents focus cues that are correct or nearly correct for the simulated scene. When focus cues are correct or nearly correct,

1. the time required to identify a stereoscopic stimulus is reduced,

2. stereoacuity in a time-limited task is increased,

3. distortions in perceived depth are reduced, and

4. viewer fatigue and discomfort are reduced.

These results have important implications for vision research and the design and use of displays.

Acknowledgments

References

Akeley K, Watt SJ, Girshick AR, Banks MS. A stereo display prototype with multiple focal distances. ACM Transactions on Graphics 2004;23:804–813.

Backus BT, Banks MS. Estimator reliability and distance scaling in stereoscopic slant perception. Perception 1999;28:217–242. [PubMed: 10615462]

Backus BT, Banks MS, van Ee R, Crowell JA. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. Vision Research 1999;39:1143–1170. [PubMed: 10343832]

Baird JC, Biersdorf WR. Quantitative functions for size and distance judgments. Perception & Psychophysics 1967;2:161–166.

Banks MS, Gepshtein S, Landy MS. Why is spatial stereoresolution so low? Journal of Neuro-science 2004;24:2077–2089.

Bereby-Meyer Y, Leiser D, Meyer J. Perception of artificial stereoscopic stimuli from an incorrect viewing point. Perception & Psychophysics 1999;61:1555–1563. [PubMed: 10598469]

Blakemore C. The range and scope of binocular depth discrimination in man. The Journal of Physiology 1970;211:599–622. [PubMed: 5501054]

Braunstein ML, Liter JC, Tittle JS. Recovering three-dimensional shape from perspective translations and orthographic rotations. Journal of Experimental Psychology: Human Perception and Performance 1993;19:598–614. [PubMed: 8331315]

Buckley D, Frisby JP. Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. Vision Research 1993;33:919–933. [PubMed: 8506635]

Burge, J.; Peterson, MA.; Palmer, SE. Ordinal configural cues combine with metric disparity in depth perception; Journal of Vision. 2005. p. 5p. 534-542.http://journalofvision.org/5/6/5/

Campbell FW. The depth of field of the human eye. Journal of Modern Optics 1957;4:157–164.

Caudek C, Proffitt DR. Depth perception in motion parallax and stereokinesis. Journal of Experimental Psychology: Human Perception and Performance 1993;19:32–47. [PubMed: 8440987]

Charman WN, Whitefoot H. Pupil diameter and depth-of-field of human eye as measured by laser speckle. Optica Acta 1977;24:1211–1216.

Chen L, Kruger PB, Hofer H, Singer B, Williams DR. Accommodation with higher-order monochromatic aberrations corrected with adaptive optics. Journal of the Optical Society of America A, Optics, Image Science, and Vision 2006;23:1–8.

Cheng, H.; Barnett, JK.; Vilupuru, AS.; Marsack, JD.; Kasthurirangan, S.; Applegate, RA., et al. A population study on changes in wave aberrations with accommodation; Journal of Vision. 2004. p. 3p. 272-280.http://journalofvision.org/4/4/3/

Cochran WG. Problems arising in the analysis of a series of similar experiments. Supplement to the Journal of the Royal Statistical Society 1937;4:102–118.

Creem-Regehr SH, Willemsen P, Gooch AA, Thompson WB. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. Perception 2005;34:191–204. [PubMed: 15832569]

Cumming BG, Judge SJ. Disparity-induced and blur-induced convergence eye-movement and accommodation in the monkey. Journal of Neurophysiology 1986;55:896–914. [PubMed: 3711971]

Domini F, Caudek C. Perceiving surface slant from deformation of optic flow. Journal of Experimental Psychology: Human Perception and Performance 1999;25:426–444. [PubMed: 10205861]

Ellis, SR.; Smith, S.; Grunwald, A.; McGreevy, MW. Pictorial communication in virtual and real environments. 2. Bristol, PA: Taylor & Francis, Inc; 1993. Direction judgement error in computer generated displays and actual scenes; p. 504-526.

Emoto M, Niida T, Okano F. Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. Journal of Display Technology 2005;1:328–340.

Favalora GE, Napoli J, Hall DM, Dorval RK, Giovinco MG, Richmond MJ, et al. 100 million-voxel volumetric display. Proceedings of SPIE 2002;712:300–312.

Fincham EF. The accommodation reflex and its stimulus. British Journal of Ophthalmology 1951;35:381–393. [PubMed: 14848436]

Fincham EF, Walton J. The reciprocal actions of accommodation and convergence. The Journal of Physiology 1957;137:488–508. [PubMed: 13463783]

Frisby JP, Buckley D, Duke PA. Evidence for good recovery of lengths of real objects seen with natural stereo viewing. Perception 1996;25:129–154. [PubMed: 8733143]

Frisby JP, Buckley D, Wishart KA, Porrill J, Garding J, Mayhew JE. Interaction of stereo and texture cues in the perception of 3-dimensional steps. Vision Research 1995;35:1463–1472. [PubMed: 7645275]

Fry GA. Further experiments on the accommodation convergence relationship. American Journal of Optometry 1939;16:125.

Fry GA, Kent PR. The effects of base-in and base-out prisms on stereo-acuity. American Journal of Optometry and Physiological Optics 1944;21:492–507.

Garding J, Porrill J, Mayhew JE, Frisby JP. Stereopsis, vertical disparity and relief transformations. Vision Research 1995;35:703–722. [PubMed: 7900308]

Ghahramani, Z.; Wolpert, DM.; Jordan, MI. Computational models of sensorimotor integration. In: Morasso, PG.; Sanguineti, V., editors. Self-organization, computational maps, and motor control. Amsterdam: North-Holland; 1997. p. 117-147.

Granger EM, Cupery KN. Optical merit function (SQF), which correlates with subjective image judgments. Photographic Science and Engineering 1972;16:221–230.

Green DG, Powers MK, Banks MS. Depth of focus, eye size and visual acuity. Vision Research 1980;20:827–835. [PubMed: 7467137]

Häkkinen, J.; Pölönen, M.; Takatalo, J.; Nyman, G. Simulator sickness in virtual display gaming: A comparison of stereoscopic and non-stereoscopic situations. Proceedings of the 8th Conference on Human–Computer Interaction With Mobile Devices and Services; New York: ACM; 2006. p. 227-230.

Hillis JM, Banks MS. Are corresponding points fixed? Vision Research 2001;41:2457–2473. [PubMed: 11483177]

Hillis, JM.; Watt, SJ.; Landy, MS.; Banks, MS. Slant from texture and disparity cues: Optimal cue combination; Journal of Vision. 2004. p. 1p. 967-992.http://journalofvision.org/4/12/1/

Hogervorst MA. Biases in three-dimensional structure-from-motion arise from noise in the early visual system. Proceedings of the Royal Society B: Biological Sciences 1998;265:1587–1593.

Hogervorst MA, Eagle RA. The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. Journal of Experimental Psychology: Human Perception and Performance 2000;26:934–955. [PubMed: 10884003]

Howard, IP.; Rogers, BJ. Seeing in depth. Toronto: University of Toronto Press; 2002.

Howarth PA, Costello PJ. The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. Displays 1997;18:107–116.

Jacobs RA. Optimal integration of texture and motion cues to depth. Vision Research 1999;39:3621–3629. [PubMed: 10746132]

Jin ZX, Zhang YJ, Wang X, Plocher T. Evaluating the usability of an auto-stereoscopic display. Lecture Notes in Computer Science 2007;4551:605.

Johnston EB. Systematic distortions of shape from stereopsis. Vision Research 1991;31:1351–1360. [PubMed: 1891823]

Julesz, B. Foundations of cyclopean perception. Chicago: University of Chicago Press; 1971.

Khosroyani M, Hung GK. A dual-mode dynamic model of the human accommodation system. Bulletin of Mathematical Biology 2002;64:285–299. [PubMed: 11926118]

Knill DC, Saunders JA. Do humans optimally integrate stereo and texture information for judgments of surface slant? Vision Research 2003;43:2539–2558. [PubMed: 13129541]

Kooi FL, Toet A. Visual comfort of binocular and 3D displays. Displays 2004;25:99–108.

Krishnan VV, Shirachi D, Stark L. Dynamic measures of vergence accommodation. American Journal of Optometry and Physiological Optics 1977;54:470–473. [PubMed: 931013]

Kuribayashi H, Ishigure Y, Suyama S, Takada H, Date M, Ishikawa K, et al. Effect on depth perception by a blur in a depth-fused 3-D display. Journal of the Institute of Image Information and Television Engineers 2006;60:431–438.

Loomis JM, Da Silva JA, Fujita N, Fukusima SS. Visual space perception and visually directed action. Journal of Experimental Psychology: Human Perception and Performance 1992;18:906–921. [PubMed: 1431754]

Loomis, JM.; Eby, DW. Proceedings from the workshop on visual motion. Washington, DC: Computer Society of the IEEE; 1989. Relative motion parallax and the perception of structure from motion; p. 204-211.

Lucente M. Interactive three-dimensional holographic displays: Seeing the future in depth. ACM SIGGRAPH Computer Graphics 1997;31:63–67.

Marshall JA, Burbeck CA, Ariely D, Rolland JP, Martin KE. Occlusion edge blur: A cue to relative visual depth. Journal of the Optical Society of America A, Optics, Image Science, and Vision 1996;13:681–688.

Martens TG, Ogle KN. Observations on accommodative convergence; especially its nonlinear relationships. American Journal of Ophthalmology 1959;47:455–463. [PubMed: 13617373]

Mather G. The use of image blur as a depth cue. Perception 1997;26:1147–1158. [PubMed: 9509149]

Mather G, Smith DR. Depth cue integration: Stereopsis and image blur. Vision Research 2000;40:3501–3506. [PubMed: 11115677]

Mather G, Smith DR. Blur discrimination and its relation to blur-mediated depth perception. Perception 2002;31:1211–1219. [PubMed: 12430948]

Mathews S, Kruger PB. Spatiotemporal transfer function of human accommodation. Vision Research 1994;34:1965–1980. [PubMed: 7941397]

McQuaide, SC. PhD thesis. University of Washington; 2002. Three-dimensional virtual retinal display using a deformable membrane mirror.

Menozzi M. Visual ergonomics of head-mounted displays. Japanese Psychological Research 2000;42:213–221.

Miles FA, Judge SJ, Optican LM. Optically induced changes in the couplings between vergence and accommodation. Journal of Neuro-science 1987;7:2576–2589.

Mon-Williams M, Tresilian JR. Ordinal depth information from accommodation? Ergonomics 2000;43:391–404. [PubMed: 10755661]

Mon-Williams M, Tresilian JR, Roberts A. Vergence provides veridical depth perception from horizontal retinal image disparities. Experimental Brain Research 2000;133:407–413.

Morgan MW. The clinical aspects of accommodation and convergence. American Journal of Optometry and Archives of American Academy of Optometry 1944;21:183–195.

Odom JV, Chao GM, Leys M. Symmetrical refractive error elevates stereo thresholds. Investigative Ophthalmology & Visual Science 1992;33:1375–1375.

Ogle KN. An analytical treatment of the longitudinal horopter, its measurement and application to related phenomena, especially to the relative size and shape of the ocular images. Journal of the Optical Society of America 1932;22:665–728.

Okada Y, Ukai K, Wolffsohn JS, Gilmartin B, Iijima A, Bando T. Target spatial frequency determines the response to conflicting defocus- and convergence-driven accommodative stimuli. Vision Research 2005;46:475–484. [PubMed: 16198392]

O'Shea RP, Govan DG, Sekuler R. Blur and contrast as pictorial depth cues. Perception 1997;26:599–612. [PubMed: 9488884]

Owens DA. A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings. Vision Research 1980;20:159–167. [PubMed: 7434577]

Paap KR, Ebenholtz SM. Perceptual consequences of potentiation in the extraocular muscles: Alternative explanation for adaptation to wedge prisms. Journal of Experimental Psychology: Human Perception and Performance 1976;2:457–468. [PubMed: 1010996]

Percival, AS. The prescribing of spectacles. Bristol: J. Wright; 1920.

Phillips, SR. PhD thesis. University of California; Berkeley: 1974. Ocular neurological control systems: Accommodation and the near response triad.

Rogers BJ, Bradshaw MF. Binocular judgments of depth, size, shape and absolute distance—Is the same d used for all judgments. Investigative Ophthalmology & Visual Science 1995;36:S230–S230.

Rolland JP, Krueger MW, Goon AA. Dynamic focusing in head-mounted displays. Proceedings of SPIE 1999;3639:463–470.

Sahm CS, Creem-Regehr SH, Thompson WB, Willemsen P. Throwing versus walking as indicators of distance perception in similar real and virtual environments. ACM Transactions on Applied Perception (TAP) 2005;2:35–45.

Schor C, Wood I, Ogawa J. Binocular sensory fusion is limited by spatial resolution. Vision Research 1984;24:661–665. [PubMed: 6464360]

Schor CM, Tsuetaki TK. Fatigue of accommodation and vergence modifies their mutual interactions. Investigative Ophthalmology & Visual Science 1987;28:1250–1259. [PubMed: 3610543]

Schowengerdt BT, Seibel EJ. True 3-D scanned voxel displays using single or multiple light sources. Journal of the Society for Information Display 2006;14:135–143.

Semmlow J, Wetzel P. Dynamic contributions of the components of binocular vergence. Journal of the Optical Society of America 1979;69:639–645. [PubMed: 479966]

Sheedy J, Bergstrom N. Performance and comfort on near-eye computer displays. Optometry and Vision Science 2002;79:306–312. [PubMed: 12035988]

Sullivan A. DepthCube solid-state 3D volumetric display. Proceedings of SPIE 2004;5291:279.

Suyama S, Takada H, Uehira K, Sakai S. A new method for protruding apparent 3-D images in the DFD (depth-fused 3-D) display. 2001 International Symposium Digest of Technical Papers 2001;32:1300–1303.

Takada H. The progress of high presence and 3D display technology. The depth-fused 3-D display for the eye sweetly. Optical and Electro-Optical Engineering Contact 2006;44:316–323.

Takaki, Y. Novel 3D display using an array of LCD panels. In: Liang-Chi, C., editor. Liquid crystal materials, devices, and applications IX; Proceedings of SPIE, 5003; 2003. p. 1-8 .

Thibos LN, Hong X, Bradley A, Cheng X. Statistical variation of aberration structure and image quality in a normal population of healthy eyes. Journal of the Optical Society of America A, Optics, Image Science, and Vision 2002;19:2329–2348.

Todd JT. The visual perception of 3D shape. Trends in Cognitive Sciences 2004;8:115–121. [PubMed: 15301751]

Todd JT, Bressan P. The perception of 3-dimensional affine structure from minimal apparent motion sequences. Perception & Psychophysics 1990;48:419–430. [PubMed: 2247324]

Torii M, Okada Y, Ukai K, Wolffsohn JS, Gilmartin B. Dynamic measurement of accommodative responses while viewing stereoscopic images. Journal of Modern Optics 2008;55:557–567.

Tucker J, Charman WN, Ward PA. Modulation dependence of the accommodation response to sinusoidal gratings. Vision Research 1986;26:1693–1707. [PubMed: 3617510]

Ukai, K. Visual fatigue caused by viewing stereoscopic images and mechanism of accommodation. Proceedings of the First International Symposium on University Communication; 2007. p. 176-179.

Vishwanath D, Girshick AR, Banks MS. Why pictures look right when viewed from the wrong place. Nature Neuroscience 2005;8:1401–1410.

Wallach H, Norris CM. Accommodation as a distance-cue. American Journal of Psychology 1963;76:659–664. [PubMed: 14082663]

Walsh G, Charman WN. Visual sensitivity to temporal change in focus and its relevance to the accommodation response. Vision Research 1988;28:1207–1221. [PubMed: 3253992]

Wann JP, Mon-Williams M. Health issues with virtual reality displays: What we do know and what we don't. ACM SIGGRAPH Computer Graphics 1997;31:53–57.

Wann, JP.; Mon-Williams, M. Measurement of visual aftereffects following virtual environment exposure. In: Stanney, KM., editor. Handbook of virtual environments: Design, implementation, and applications. Hillsdale, NJ: Lawrence Erlbaum Associates; 2002. p. 731-749.

Ward PA. The effect of spatial frequency on steady-state accommodation. Ophthalmic & Physiological Optics 1987;7:211–217. [PubMed: 3684273]

Watt, SJ.; Akeley, K.; Ernst, MO.; Banks, MS. Focus cues affect perceived depth; Journal of Vision. 2005. p. 7p. 834-862.http://journalofvision.org/5/10/7/

Westheimer G, McKee SP. Stereoscopic acuity with defocused and spatially filtered retinal images. Journal of the Optical Society of America 1980;70:772–778.

Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. Perception & Psychophysics 2001a;63:1293–1313. [PubMed: 11800458]

Wichmann FA, Hill NJ. The psychometric function: II. Bootstrap-based confidence intervals and sampling. Perception & Psychophysics 2001b;63:1314–1329. [PubMed: 11800459]

Willemsen P, Gooch AA, Thompson WB, Creem-Regehr SH. Effects of stereo viewing conditions on distance perception in virtual environments. 2007 Manuscript submitted for publication.

Wood ICJ. Stereopsis with spatially-degraded images. Ophthalmic & Physiological Optics 1983;3:337–340. [PubMed: 6646769]

Yamazaki T, Kamijo K, Fukuzumi S. Quantitative evaluation of visual fatigue encountered in viewing stereoscopic 3D displays: Near-point distance and visual evoked potential study. Proceedings of the Society for Information Display 1990;31:245–247.

Yano S, Emoto M, Mitsuhashi T. Two factors in visual fatigue caused by stereoscopic HDTV images. Displays 2004;25:141–150.

Yano S, Ide S, Mitsuhashi T, Thwaites H. A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. Displays 2002;23:191–201.
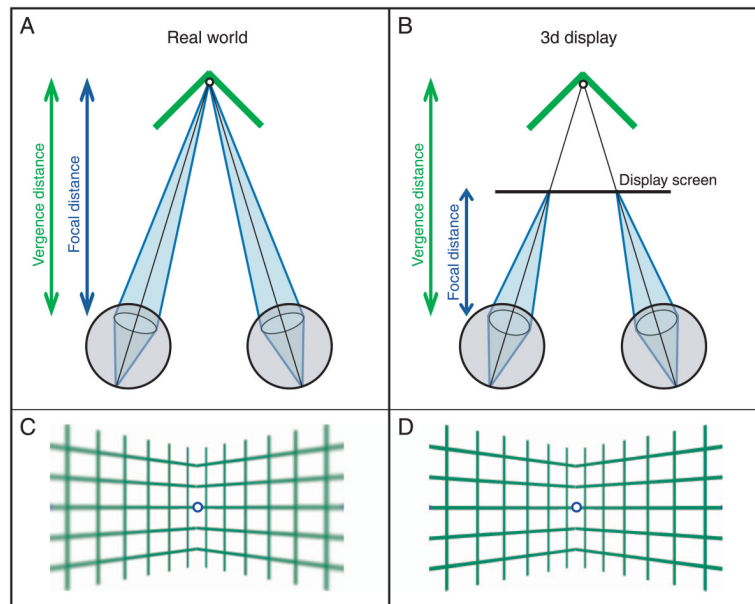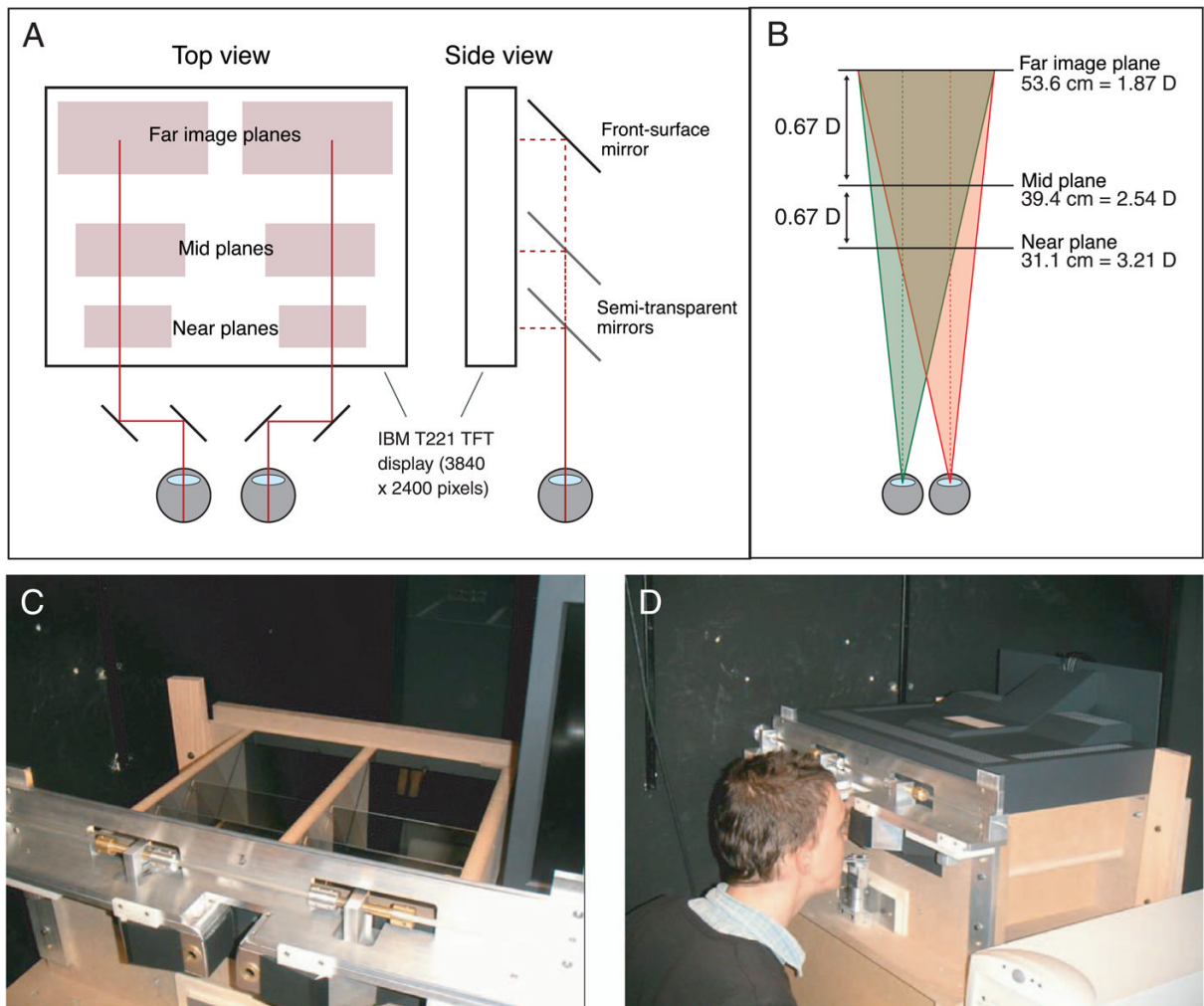
Figure 1.
Vergence and focal distance with real stimuli and stimuli presented on conventional 3D displays. (A) The viewer is fixated and focused on the vertex of a hinge. Vergence distance is the distance to the vertex. Vergence response is the distance to the intersection of the eyes' lines of sight. Focal distance is the distance to which the eye would have to focus to create a sharply focused image. Accommodative response is the distance to which the eye is accommodated. (B) The viewer is fixated on the simulated hinge vertex on a computer display screen. Vergence distance is the same as in panel A. Focal distance is now the distance to the display. (C) The appearance of the stimulus when the viewer is accommodated to the vertex of a real hinge. In the retinal image, the joined planes (the sides) of the hinge are blurred relative to the vertex. (D) Appearance when the viewer is accommodated to the vertex of a simulated hinge. The sides and the vertex are equally sharp in the retinal image.
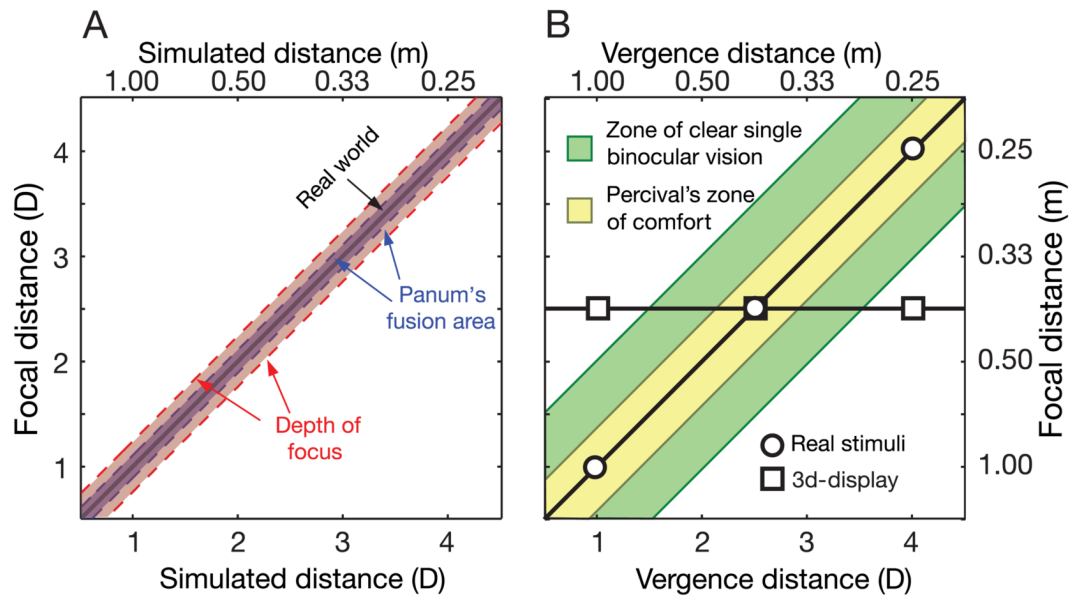
Figure 2.
Consequences of vergence–accommodation coupling. The panels plot focal distance as a function of simulated (or vergence) distance. The bottom abscissa and the left ordinate have units of diopters. The top abscissa and the right ordinate have the corresponding values in meters. (A) Depth of focus and Panum's fusion area in diopters. Here we simulate real objects, so the vergence and the focal distances specified by the object are equal to one another. The diagonal line represents the viewer's vergence and accommodation assuming accurate responses to the object. Vertical cross sections of the red region represent the eye's depth of focus. Horizontal cross sections of the blue region represent Panum's fusion area (15 arcmin). Note that the sizes of the zones of clear vision and of single vision remain constant in diopters. Because of this, we will express simulated and vergence distances in diopters instead of in meter-angles, which are the conventional units in optometry and ophthalmology. (B) Zones of clear single vision and of comfort. The green area represents the zone of clear single binocular vision: the range of vergence and accommodative responses that young adults can achieve without excessive effort and without exceeding depth of focus or Panum's area. The yellow area represents Percival's zone of comfort: the range of responses viewers can achieve without discomfort. Circles represent three real-world stimuli; squares represent the corresponding stimuli on a conventional 3D display.

Figure 3.
Fixed-viewpoint, volumetric display. The IBM T221 LCD panel is viewed through semi-transparent and front-surface mirrors such that each eye sees three superimposed images. (A) Schematic. The left side is a top view. The T221 display is placed "face down" on top of the apparatus. Shaded rectangles represent the portions of the display surface that are the far, mid, and near image planes. Red lines represent left and right eyes' lines of sight if the eyes are in parallel gaze (vergence = 0°). Periscope optics delivers the images to the eyes. The right side of Panel A is a side view. It shows the arrangement of the mirrors for creating each eye's view. The images from the three planes travel different distances before arriving at the eyes. The far image plane is reflected off the front-surface mirror and transmitted through two semi-transparent mirrors. The near image plane is reflected off one semi-transparent mirror. (B) The focal distances of the image planes and the viewing frusta for the two eyes. The focal distances of the image planes are 31.1 (near), 39.4 (mid), and 53.6 cm (far) (3.21, 2.54, and 1.87 D; 0.67 D separations). The viewing frusta were skewed so that the binocular overlap was maximized at the far image plane. (C) The display is removed from the top to expose the mirrors. The apertures for viewing the stimuli are in the lower left. A worm gear allows the aperture separation to be adjusted to the observer's inter-ocular distance. (D) An observer viewing stimuli with the display in place. Details in Akeley et al. (2004).
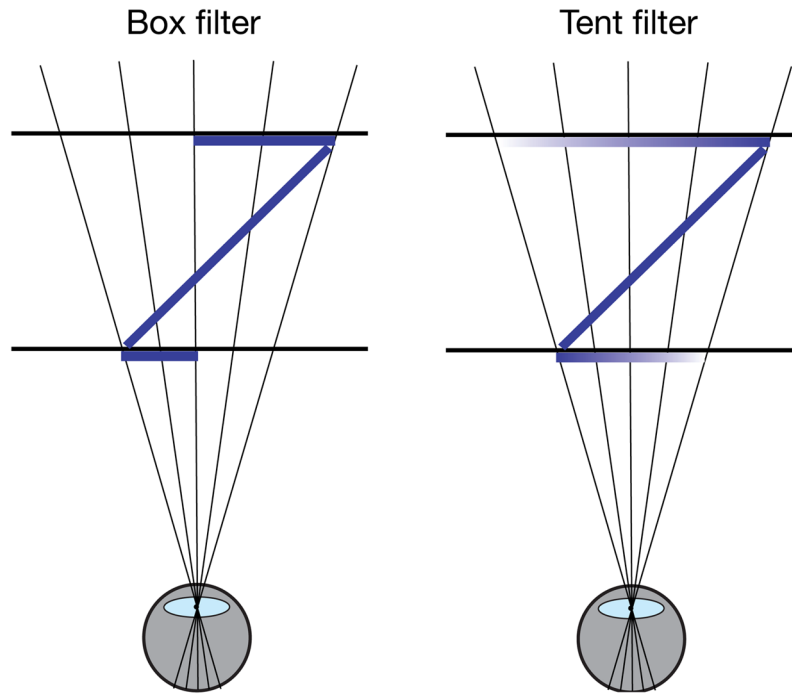
Figure 4.
Pixel lighting with box and tent depth filters. The horizontal lines represent two image planes, the thick blue diagonal line the surface we wish to draw, and the thick blue horizontal lines the pixel intensities on the image planes; dark blue represents high intensity and pale blue represents low intensity. The box filter illuminates the far image plane at high intensity for all regions in which the simulated surface is closer to the far than the near plane. It illuminates the near plane at high intensity for all regions in which the simulated surface is closer to that plane. The tent filter weights intensities according to the proportion of the dioptric distance between the two image planes. This occurs along each line of sight. The sum of the two intensities is the same as the intensity on one plane in a conventional display. For details, see Akeley et al. (2004).
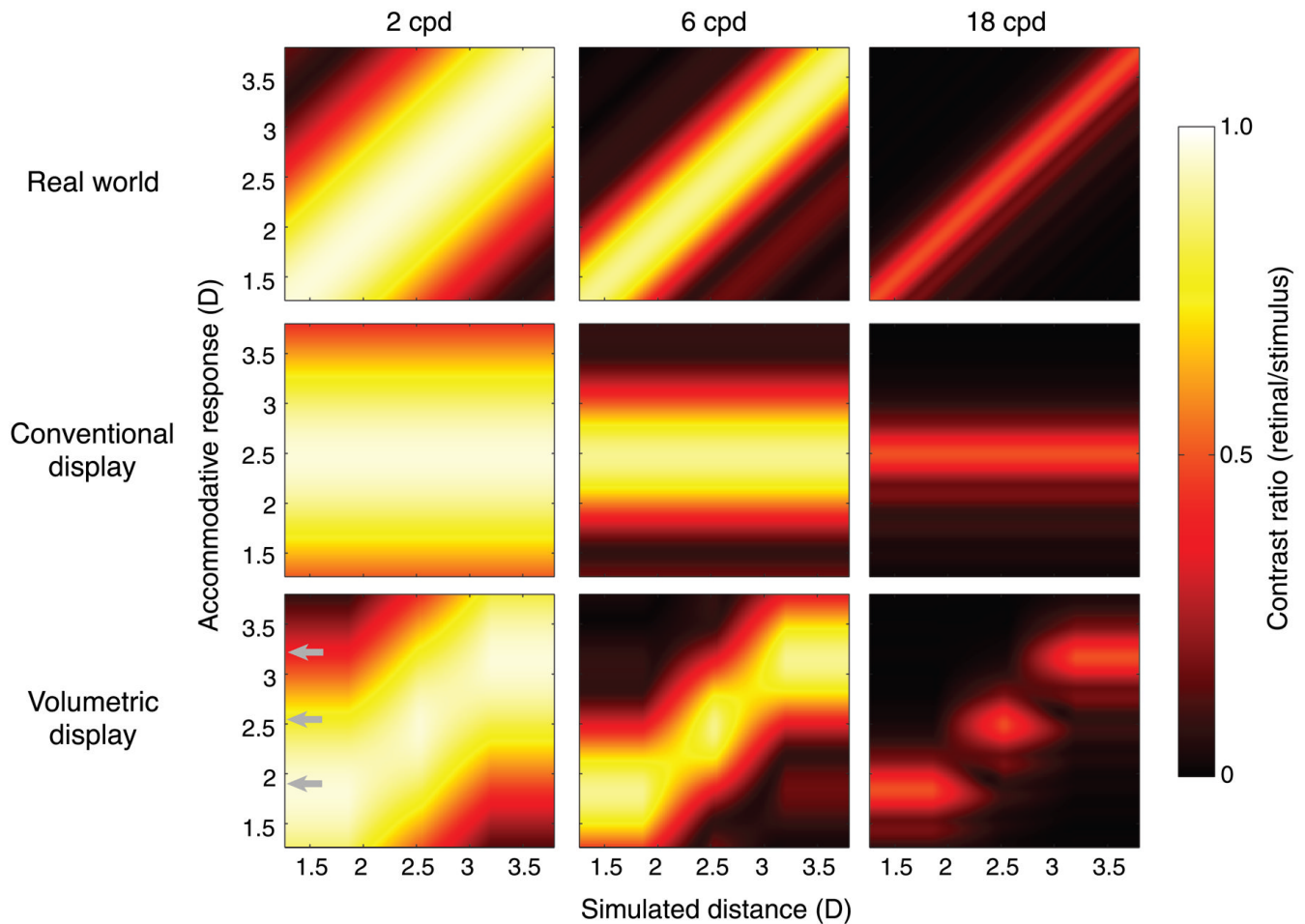
Figure 5.
Retinal-image contrast for different displays and hypothetical accommodative responses. The top, middle, and bottom rows represent stimuli in the real world, in a conventional 3D display, and in our multi-plane display, respectively. The left, middle, and right columns represent spatial sinusoids of 2, 6, and 18 cpd, respectively. The abscissas are the actual or the simulated distance to the stimulus in diopters. The ordinates are the accommodative response in diopters. The colors represent contrast ratio: retinal-image contrast divided by stimulus contrast. The panels were generated based on the optical aberrations of the left eye of author DMH for a 4.5-mm pupil. His eye is emmetropic, so the maximum retinal contrast in the upper row occurs when the accommodative response is very close to the stimulus distance. The distance to the conventional 3D display was set at 2.5 D (40 cm). The secondary bands in the 6- and 18-cpd panels are caused by the ringing in the MTF of the defocused eye. The distances to the three planes in the multi-plane display were 1.87, 2.54, and 3.21 D, a range of 1.33 D. Wavefront measurements and analysis tools were provided by Austin Roorda, UC Berkeley.
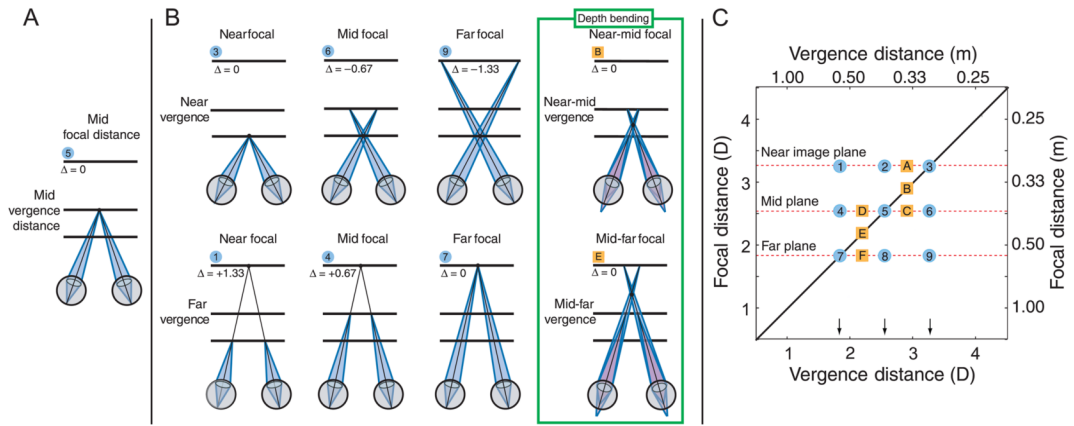
Figure 6.
The conditions in Experiments 1 and 2. The horizontal lines represent the image planes in the apparatus. The thin black lines represent the visual axes of the eyes; those axes intersect at the vergence distance. The light blue regions represent the focal distance. Panel A represents the fixation stimulus before the presentation of the test stimulus. Δ is the difference in diopters between the vergence and the focal distances. Panel B depicts eight of the test stimuli in the experiments. For the stimuli marked 3, 7, B, and E, the focal and the vergence distances were equal to one another (cues-consistent). The rightmost column, outlined in green, depicts the two stimuli with depth-weighted blending (see General methods section). In these stimuli, the vergence distance was the dioptric midpoint between two image planes; 50% of the light came from each of those planes. Panel C represents all the stimuli in the experiments. The horizontal dashed lines represent the three image planes. The vertical arrows represent the vergence distances. Stimuli at positions 1, 3, 4, 5, 6, 7, 9, A, B, C, D, E, and F were presented in Experiments 1 and 2; in B and E, the focal stimuli are depth-weighted blends. Experiment 3 used stimuli at 1–9. Experiment 4 used stimuli at 4, 5, and 6 for the cues-inconsistent session, and stimuli at 3, 5, and 7 in the cues-consistent session.
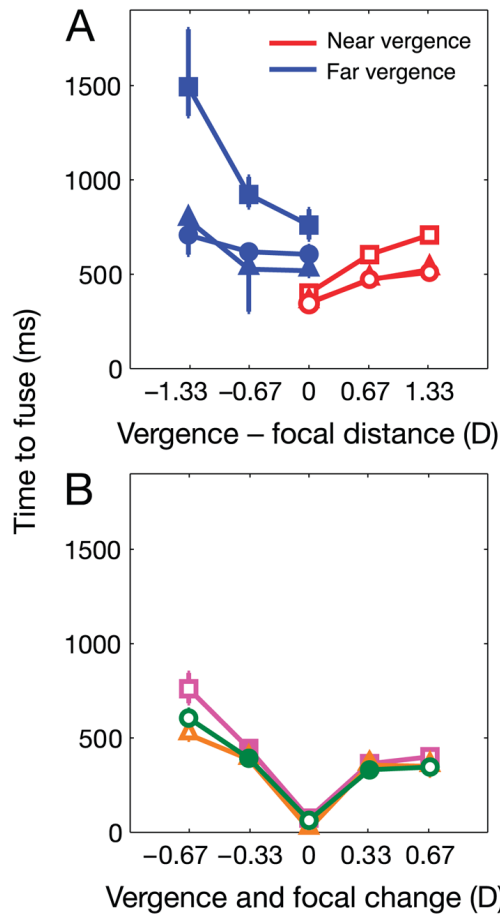
Figure 7.
Results of Experiment 1. (A) Results for the non-blended stimuli. The abscissa represents the difference between the vergence and the focal distances (vergence minus focal) in diopters. The ordinate represents the stimulus duration required for the observer to correctly identify the orientation of the cyclopean stimulus 75% of the time. The data for the ±0.33 D conflicts are not shown. Different symbols represent the data from different observers: squares for ARG, triangles for DS, and circles for BGS. The blue filled symbols represent data when the vergence distance jumped from the mid to the far image plane; red unfilled symbols represent data when the vergence distance jumped from the mid to the near plane. Error bars are 95% confidence intervals. (B) Results for the cues-consistent stimuli. The abscissa represents the change in the vergence and the focal distance from the fixation to the test stimulus; positive values are changes in which the distance decreased and negative values are those in which the distance increased. The ordinate represents the stimulus duration required to correctly identify stimulus orientation 75% of the time. Different symbols represent data from different observers: purple squares for ARG, orange triangles for DS, and green circles for BGS. The unfilled symbols represent data with non-blended stimuli in which the focal and the vergence stimuli were at one of the image planes. The filled symbols are data with blended stimuli in which focal and vergence distance were between image planes. Error bars are 95% confidence intervals.
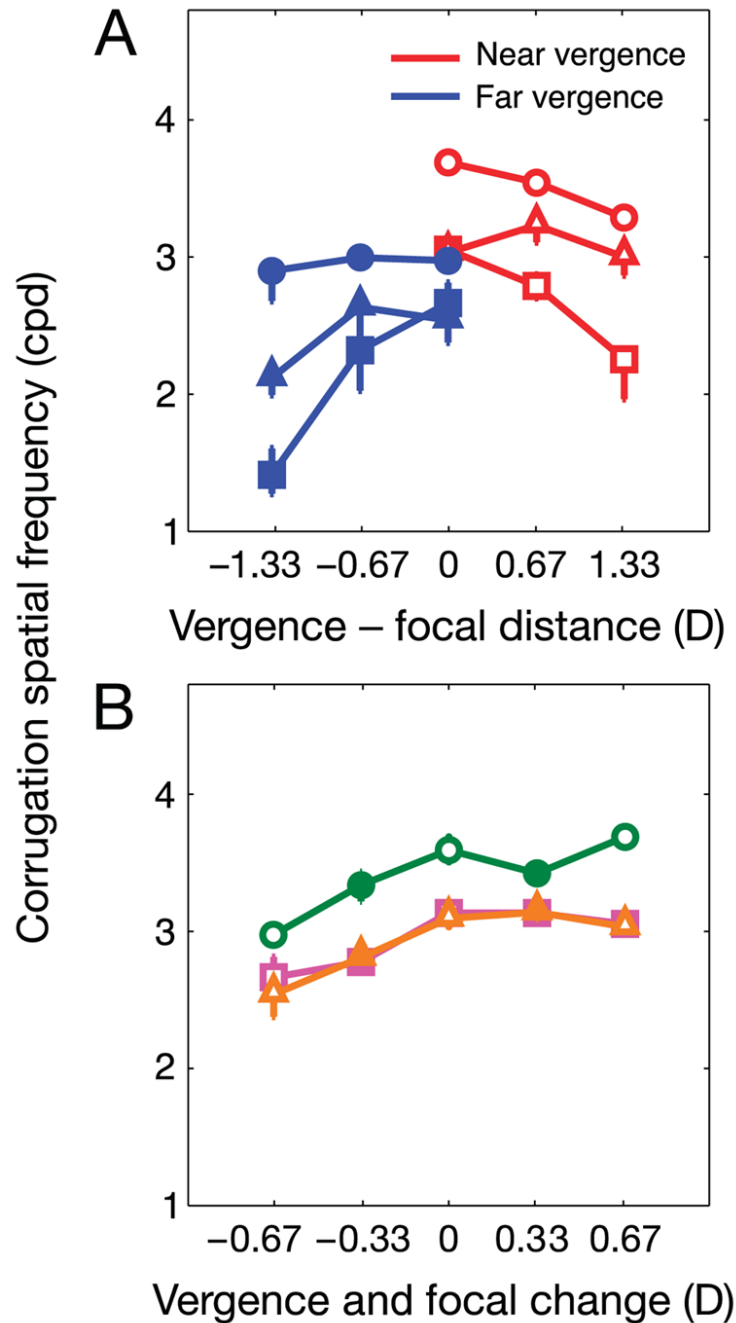
Figure 8.
Results from Experiment 2. The format is the same as Figure 7 except the ordinate represents the highest spatial frequency at which observers could identify the corrugation orientation correctly 75% of the time. (A) Results for non-blended stimuli with near or far vergence distances. The abscissa represents the conflict between the vergence and the focal distances in diopters. The data for the ±0.33 D conflicts are not shown. Different symbol types represent the data from different observers: squares for ARG, triangles for DS, and circles for BGS. The blue filled symbols represent the data for far vergence distances and the red unfilled symbols the data for near vergence distances. Error bars represent 95% confidence intervals. (B) Results for cues-consistent stimuli. The abscissa represents the change in the vergence and the focal

distance from the fixation to the test stimulus; positive values are changes in which the distance decreased and negative values are those in which the distance increased. Purple squares are data for ARG, orange triangles for DS, and green circles for BGS. The unfilled symbols are data with non-blended stimuli presented at an image plane, and the filled symbols are data with blended stimuli presented between image planes. Error bars are 95% confidence intervals.
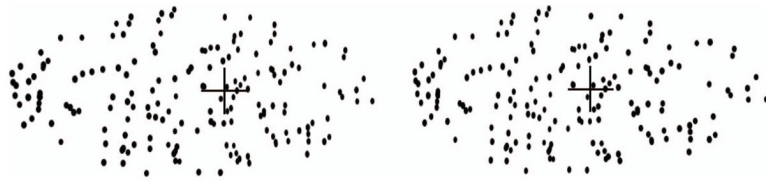
Figure 9.
The stimulus in Experiment 3. A random-dot stereogram depicting a vertical hinge in an open-book configuration. Observers indicated whether the perceived hinge angle was greater or less than 90 deg. The task could not be performed monocularly. The reader can see the 3D hinge by cross-fusing the stimulus. The experiment used a red stimulus on a black background but was reproduced here as black on white for clarity. The hinge is shown in plan view in Figure 1A.
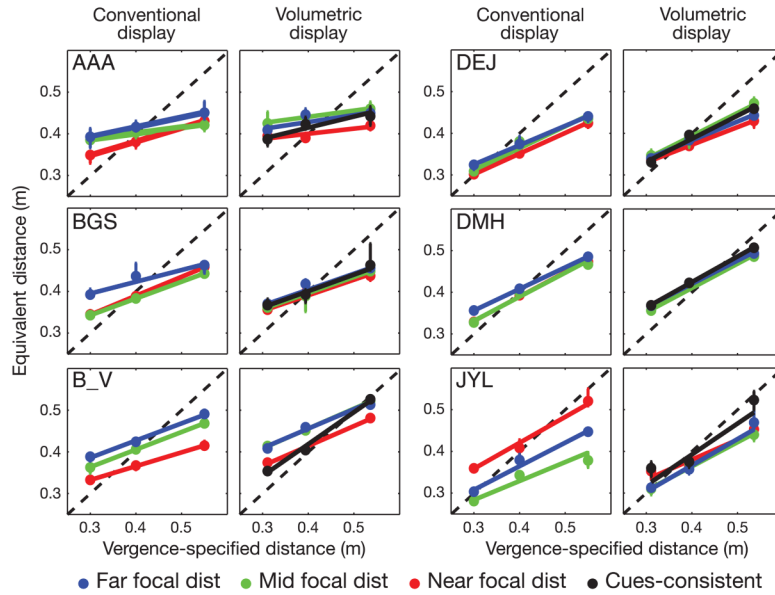
Figure 10.
Results from Experiment 3. Equivalent distance-the distance at which the disparity setting would correspond to a right angle-is plotted as a function of vergence distance. The left columns on each half of the figure show the data obtained with the conventional display. The right columns show the data obtained with the volumetric display. Each row on each half of the figure shows the data from a different observer. The abscissa in each plot is the vergence distance. The ordinate is the equivalent distance (Equation 5). The diagonal dashed lines represent the predicted data if equivalent distance were based solely on the vergence-specified distance. The red, green, and blue points represent the cues-inconsistent data from the near, mid, and far focal distances, respectively (note that 1/3 of those points are in fact cues-consistent). The colored lines are regression fits to those data points. Error bars are 95% confidence intervals. The black lines in the right column represent the data from the cues-consistent session.
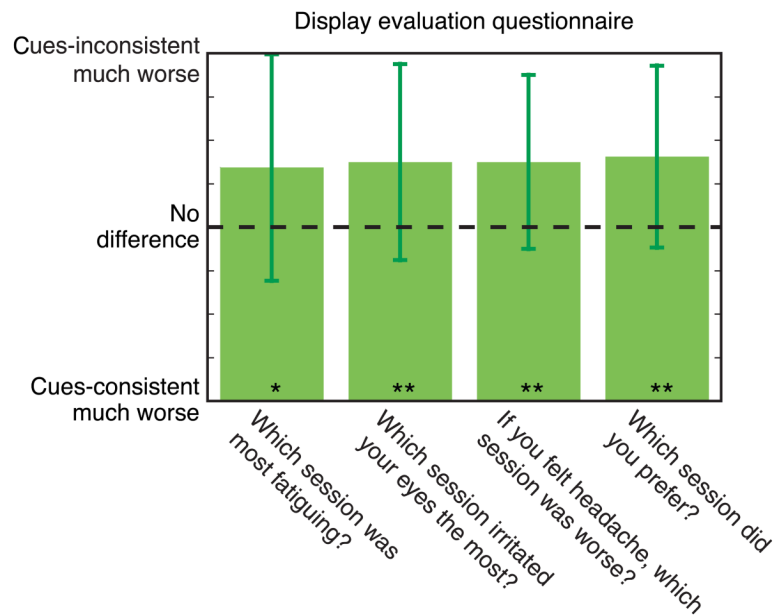
Figure 11.
Top: symptom questionnaire. Participants completed this questionnaire after each session. For each of the five questions, they indicated the severity of their symptoms at that moment. Bottom: display-evaluation questionnaire. Participants completed this questionnaire after the second session. For each of four questions, they indicated which session was better (or worse).
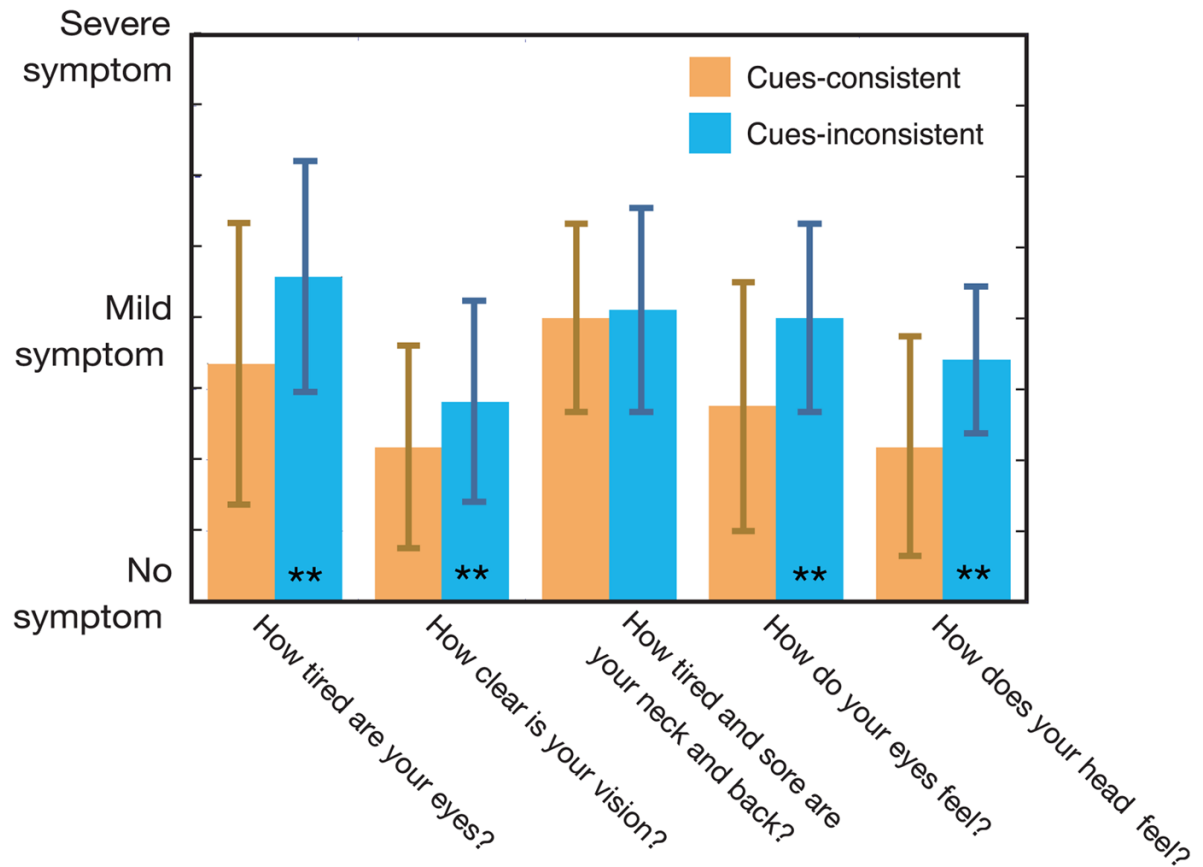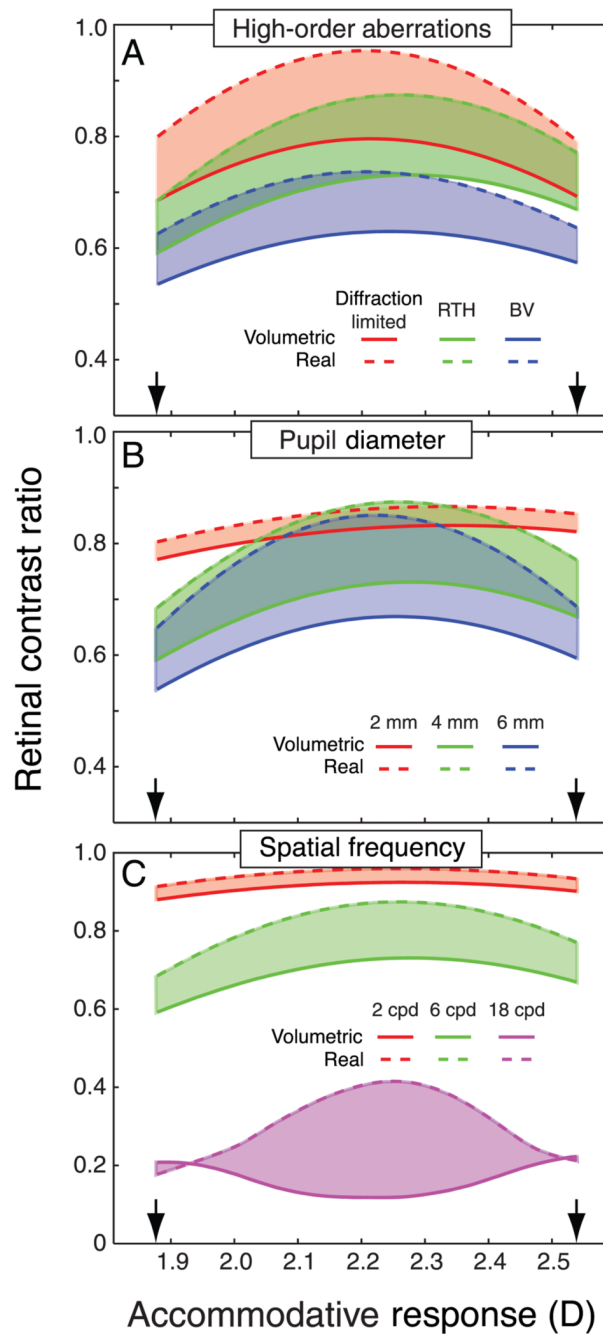
Figure 12.
Results from the symptom questionnaire (Figure 11 top). The average severity of the reported symptom is plotted for each of the five questionnaire items. Blue bars are the average reported symptoms after the cues-inconsistent session, and orange bars are the average symptoms after the cues-consistent session. Larger values are associated with more severe symptoms. Error bars represent ±1 standard deviation. This graph shows data from 11 subjects, with six of them contributing twice. The double asterisks denote significantly worse symptoms reported in the cues-inconsistent, than the cues-consistent session ($p < 0.025$, Wilcoxon signed-rank test, one tailed).

Figure 13.
Results from the display-evaluation questionnaire (Figure 11, bottom). The average comparative rating of a pair of sessions (one cues-consistent and one cues-inconsistent) is plotted for each of the four questionnaire items. The dashed horizontal line indicates a response of "no difference." Values above that line indicate more favorable responses for the cues-consistent session. Error bars represent ±1 standard deviation. The graph shows data from 11 subjects, five of them contributing twice. (One subject misunderstood the instructions, and her first display-evaluation questionnaire was omitted.) Asterisks mean that the cues-consistent rating was significantly more favorable than the cues-inconsistent: double asterisks indicate $p < 0.025$; single asterisk indicates $p < 0.05$ (Wilcoxon signed-rank test, one tailed).
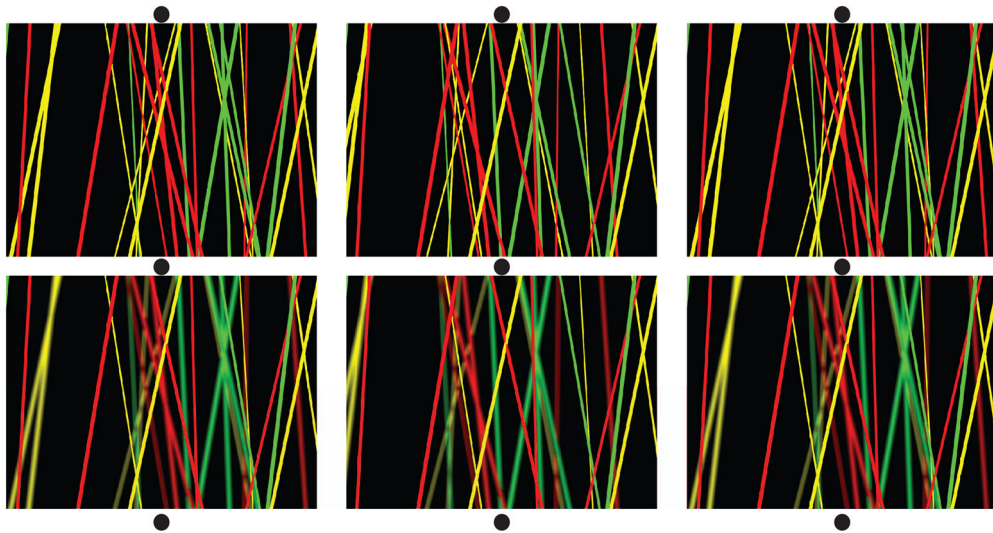
Figure 14.
Retinal images with real-world and volumetric viewing. In each panel, the abscissa is the accommodative response in diopters and the ordinate is the retinal-contrast ratio: retinal-image contrast divided by object contrast. The real-world stimulus is always presented at 2.21 D. The volumetric stimulus is always presented on image planes at 1.87 and 2.54 D (represented by the arrows) to simulate an object at 2.21 D; this mimics the situation in our volumetric display. The dashed and the solid lines represent the contrast ratios created by real and volumetric stimuli, respectively; the shaded areas represent the difference. The stimulus in every case is a vertical sinusoidal grating. (A) The effect of viewer's optical aberrations. Pupil diameter is 4 mm and stimulus spatial frequency is 6 cpd. The red, green, and blue curves represent

respectively the contrast ratios for a diffraction-limited eye, the eye of a viewer with typical aberrations, and the eye of a viewer with larger-than-typical aberrations. (B) The effect of pupil size. The viewer has typical aberrations. The stimulus is 6 cpd. The red, green, and blue curves represent the contrast ratios for pupil diameters of 2, 4, and 6 mm, respectively. (C) The effect of stimulus spatial frequency. The viewer has typical aberrations and pupil diameter is 4 mm. The red, green, and blue curves represent the contrast ratios for spatial frequencies of 2, 6, and 18 cpd, respectively.

Figure 15.
Salience of diplopia with and without blur. The stereograms depict two frontoparallel planes of sticks. In the upper stereogram, the sticks in the near and far planes are rendered sharp. In the lower stereogram, the sticks in the near plane are rendered sharp and the ones in the far plane are blurred. Cross fuse for the left pairs; divergently fuse for the right pairs.

Table 1

Results from Experiment 3. Each row shows the data for a different observer. The first column provides the observers' initials. The second indicates the experimental condition. The third shows the data from the conventional display condition, and the fourth column shows the data from the volumetric display condition. The entries in the third and fourth columns are the slopes of the linear-regression fits to the data in Figure 10. The entries in the cues-inconsistent condition are the averages of the three regression-line slopes for each observer. The entries in the cues-consistent, between-sessions condition are the fits to the cues-consistent data across the three inconsistent sessions (near, mid, and far vergence and focal distance). The entries in the cues-consistent, within-session condition are the fits to the data obtained with the volumetric display when focal and vergence distance were changed from trial to trial but were always equal to one another. The entries at the bottom of the table are the slopes averaged across observers.

Subject	Experimental condition	Conventional display	Volumetric display
AAA	Cues-inconsistent sessions	0.24	0.15
	Cues-consistent between-sessions	0.44	0.23
	Cues-consistent within-session		0.26
BGS	Cues-inconsistent sessions	0.38	0.38
	Cues-consistent between-sessions	0.45	0.45
	Cues-consistent within-session		0.4
B_V	Cues-inconsistent sessions	0.39	0.46
	Cues-consistent between-sessions	0.65	0.6
	Cues-consistent within-session		0.77
DEJ	Cues-inconsistent sessions	0.5	0.48
	Cues-consistent between-sessions	0.62	0.49
	Cues-consistent within-session		0.56
DMH	Cues-inconsistent sessions	0.55	0.57
	Cues-consistent between-sessions	0.63	0.57
	Cues-consistent within-session		0.61
JYL	Cues-inconsistent sessions	0.54	0.56
	Cues-consistent between-sessions	0.37	0.48
	Cues-consistent within-session		0.7
Average	Cues-inconsistent sessions	0.43	0.43
	Cues-consistent between-sessions	0.53	0.47
	Cues-consistent within-session		0.55