UC Berkeley

UC Berkeley Previously Published Works

Title

Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays

Permalink

https://escholarship.org/uc/item/9gn9x0rm

Journal

Proceedings of the National Academy of Sciences of the United States of America, 114(9)

ISSN

0027-8424

Authors

Padmanaban, Nitish
Konrad, Robert
Stramer, Tal
et al.

Publication Date

2017-02-28

DOI

10.1073/pnas.1617251114

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays

Nitish Padmanaban[a], Robert Konrad[a], Tal Stramer[a], Emily A. Cooper[b,1], and Gordon Wetzstein[a,1]

[a]Department of Electrical Engineering, Stanford University, Stanford, CA 94305; and [b]Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH 03755

Edited by Wilson S. Geisler, The University of Texas at Austin, Austin, TX, and approved January 6, 2017 (received for review October 19, 2016)

From the desktop to the laptop to the mobile device, personal computing platforms evolve over time. Moving forward, wearable computing is widely expected to be integral to consumer electronics and beyond. The primary interface between a wearable computer and a user is often a near-eye display. However, current generation near-eye displays suffer from multiple limitations: they are unable to provide fully natural visual cues and comfortable viewing experiences for all users. At their core, many of the issues with near-eye displays are caused by limitations in conventional optics. Current displays cannot reproduce the changes in focus that accompany natural vision, and they cannot support users with uncorrected refractive errors. With two prototype near-eye displays, we show how these issues can be overcome using display modes that adapt to the user via computational optics. By using focus-tunable lenses, mechanically actuated displays, and mobile gaze-tracking technology, these displays can be tailored to correct common refractive errors and provide natural focus cues by dynamically updating the system based on where a user looks in a virtual scene. Indeed, the opportunities afforded by recent advances in computational optics open up the possibility of creating a computing platform in which some users may experience better quality vision in the virtual world than in the real one.

virtual reality | augmented reality | 3D vision | vision correction | computational optics

Emerging virtual reality (VR) and augmented reality (AR) systems have applications that span entertainment, education, communication, training, behavioral therapy, and basic vision research. In these systems, a user primarily interacts with the virtual environment through a near-eye display. Since the invention of the stereoscope almost 180 years ago (1), significant developments have been made in display electronics and computer graphics (2), but the optical design of stereoscopic near-eye displays remains almost unchanged from the Victorian age. In front of each eye, a small physical display is placed behind a magnifying lens, creating a virtual image at some fixed distance from the viewer (Fig. 1*A*). Small differences in the images displayed to the two eyes can create a vivid perception of depth, called stereopsis.

However, this simple optical design lacks a critical aspect of 3D vision in the natural environment: changes in stereoscopic depth are also associated with changes in focus. When viewing a near-eye display, users' eyes change their vergence angle to fixate objects at a range of stereoscopic depths, but to focus on the virtual image, the crystalline lenses of the eyes must accommodate to a single fixed distance (Fig. 2*A*). For users with normal vision, this asymmetry creates an unnatural condition known as the vergence–accommodation conflict (3, 4). Symptoms associated with this conflict include double vision (diplopia), compromised visual clarity, visual discomfort, and fatigue (3, 5). Moreover, a lack of accurate focus also removes a cue that is important for depth perception (6, 7).

The vergence–accommodation conflict is clearly an important problem to solve for users with normal vision. However, how many people actually have normal vision? Correctable visual impairments caused by refractive errors, such as myopia (near-sightedness) and hyperopia (far-sightedness), affect approximately one-half of the US population (8). Additionally, essentially all people in middle age and beyond are affected by presbyopia, a decreased ability to accommodate (9). For people with these common visual impairments, the use of near-eye displays is further restricted by the fact that it is not always possible to wear optical correction.

Here, we first describe a near-eye display system with focus-tunable optics—lenses that change their focal power in real time. This system can provide correction for common refractive errors, removing the need for glasses in VR. Next, we show that the same system can also mitigate the vergence–accommodation conflict by dynamically providing near-correct focus cues at a wide range of distances. However, our study reveals that this conflict should be addressed differently depending on the age of the user. Finally, we design and assess a system that integrates a stereoscopic eye tracker to update the virtual image distance in a gaze-contingent manner, closely resembling natural viewing conditions. Compared with other focus-supporting display designs (10–18) (details are in *SI Appendix*), these adaptive technologies can be implemented in near-eye systems with readily available optoelectronic components and offer uncompromised image resolution and quality. Our results show how computational optics can increase the accessibility of VR/AR and improve the experience for all users.

Results

Near-Eye Display Systems with Adaptive Focus. In our first display system, a focus-tunable liquid lens is placed between each eye and a high-resolution microdisplay. The focus-tunable lenses allow for adaptive focus—real-time control of the distance to the virtual image of the display (Fig. 1*A*, green arrows). The lenses are driven by the same computer that controls the displayed images, allowing for precise temporal synchronization between the

Significance

Wearable displays are becoming increasingly important, but the accessibility, visual comfort, and quality of current generation devices are limited. We study optocomputational display modes and show their potential to improve experiences for users across ages and with common refractive errors. With the presented studies and technologies, we lay the foundations of next generation computational near-eye displays that can be used by everyone.

Author contributions: N.P., R.K., E.A.C., and G.W. designed research; N.P., R.K., and T.S. performed research; N.P. and E.A.C. analyzed data; and N.P., R.K., E.A.C., and G.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. Email: emily.a.cooper@dartmouth.edu or gordon.wetzstein@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617251114/-/DCSupplemental.

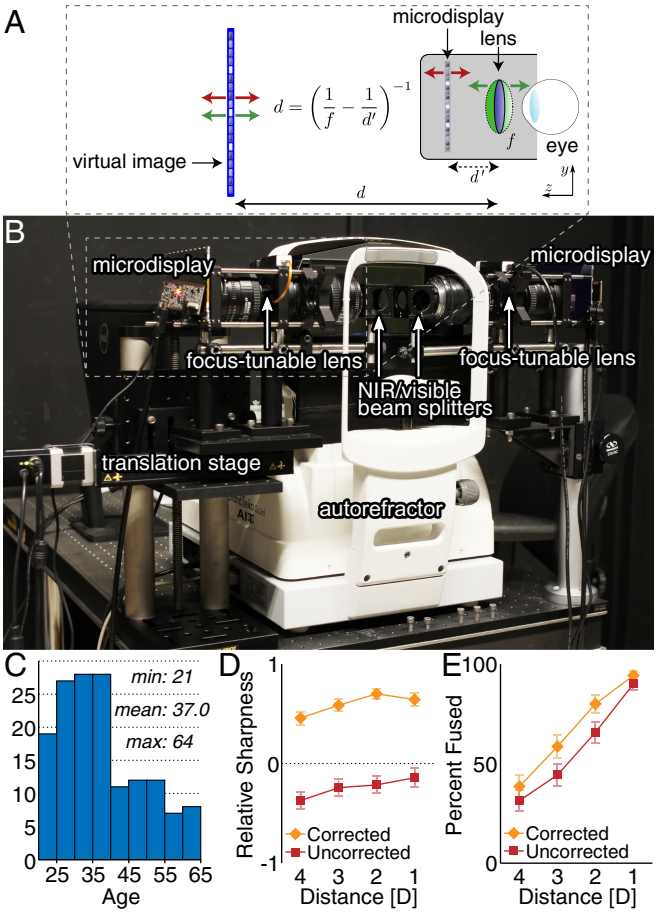

Fig. 1. (*A*) A typical near-eye display uses a fixed focus lens to show a magnified virtual image of a microdisplay to each eye (the eyes cannot accommodate at the very near microdisplay's physical distance). The focal length of the lens, *f*, and the distance to the microdisplay, *d′*, determine the distance of the virtual image, *d*. Adaptive focus can be implemented using either a focus-tunable lens (green arrows) or a fixed focus lens and a mechanically actuated display (red arrows), so that the virtual image can be moved to different distances. (*B*) A benchtop setup designed to incorporate adaptive focus via focus-tunable lenses and an autorefractor to record accommodation. A translation stage adjusts intereye separation, and NIR/visible light beam splitters allow for simultaneous stimulus presentation and accommodation measurement. (*C*) Histogram of user ages from our main studies. (*D* and *E*) The system from *B* was used to test whether common refractive errors could quickly be measured and corrected for in an adaptive focus display. Average (*D*) sharpness ratings and (*E*) fusibility for Maltese cross targets are shown for each of four distances: 1–4 D. The *x* axis is reversed to show nearer distances to the left. Targets were shown for 4 s. Red data points indicate users who did not wear refractive correction, and orange data points indicate users for whom correction was implemented on site by the tunable lenses. Values of -1, 0, and 1 correspond to responses of blurry, medium, and sharp, respectively. Error bars indicate SE across users.

virtual image distance and the onscreen content. Thus, the distance can be adjusted to match the requirements of a particular user or particular application. Details are in *SI Appendix*, and related systems are described in refs. 19–21. This system was table-mounted to allow for online measurements of the accommodative response of the eyes via an autorefractor (Fig. 1*B*), similar to the objective measurements in ref. 14, but the compact liquid lenses can fit within conventional-type head-mounted casings for VR systems. Adaptive focus can also be achieved by combining fixed focus lenses and a mechanically adjustable display (Fig. 1*A*, red arrows) (11). This approach is used for our second display system, which has the advantage of having a much larger field of view, and it will be discussed later. To assess how adaptive focus can be integrated into VR systems so as to optimize the display for the broadest set of users, we conducted a series of studies examining ocular responses and visual perception in VR. Our main user population was composed of adults with a wide range of ages (n = 153, age range = 21–64 y old) (Fig. 1*C*) and different refractive errors (79 wore glasses and 19 wore contact lenses).

Correcting Myopia and Hyperopia in VR. Before examining the vergence–accommodation conflict, we first tested whether a simple procedure can measure a user's refractive error and correct it natively in a VR system with adaptive focus. Refractive errors, such as myopia and hyperopia, are extremely common (22) and result when the eye's lens does not produce a sharp image on the retina for objects at particular distances. Although these impairments can often be corrected with contact lenses or surgery, many people wear eyeglasses. Current generation VR/AR systems require the user to wear their glasses beneath the near-eye display system. Although wearing glasses is technically possible with some systems, user reviews often cite problems with fit and comfort, which are likely to increase as the form factor of near-eye displays decreases.

Users (n = 70, ages 21–64 y old) were first tested using a recently developed portable device that uses a smartphone application to interactively determine a user's refractive error without clinician intervention, including the spherical lens power required for clear vision (NETRA; EyeNetra, Inc.) (23). After testing, each user performed several tasks in VR without wearing his/her glasses. Stimuli were presented under two conditions: uncorrected (the display's virtual image distance was 1.3 m) and corrected (the virtual image was adjusted to appear at 1.3 m after the correction was applied). Note that the tunable lenses do not correct astigmatism. We assessed the sharpness and fusibility of a Maltese cross under both conditions. The conditions were randomly interleaved along with four different stereoscopic target distances: 1–4 Diopters (D; 1.0, 0.5, 0.33, and 0.25 m, respectively). Users were then asked (*i*) how sharp the target was (blurry, medium, or sharp) and (*ii*) whether the target was fused (i.e., not diplopic).

As expected, the corrected condition substantially increased the perceived sharpness of targets at all distances (Fig. 1*D*). This condition also increased users' ability to fuse targets (Fig. 1*E*). Logistic regressions indicated significant main effects for both condition and distance. The odds ratios for correction were 4.05 [95% confidence interval (ci) = 3.25–5.05] and 1.54 (ci = 1.20–1.98) for sharpness and fusibility, respectively. The distance odds ratios were 0.77 and 0.21, respectively (all $ps \leq 0.01$), indicating reductions in both sharpness and fusibility for nearer distances.

Importantly, the VR-corrected sharpness and fusibility were comparable with those reported by people wearing their typical correction, who participated in the next study (called the conventional condition). Comparing responses between these two groups of users reveals that, across all distances, the average sharpness values for the corrected and conventional conditions were 0.60 and 0.63, respectively. The percentages fused were 68 and 74%, respectively. This result suggests that fast, user-driven vision testing can provide users with glasses-free vision in VR that is comparable with the vision that they have with their own correction.

We also assessed overall preference between the two conditions (corrected and uncorrected) in a less structured session. A target moved sinusoidally in depth within a complex virtual scene, and the user could freely toggle between conditions to select the one that was more comfortable to view; 80% of users preferred the corrected condition, which is significantly above chance (binomial probability distribution; $p \ll 0.001$). Those that preferred the uncorrected condition may have had

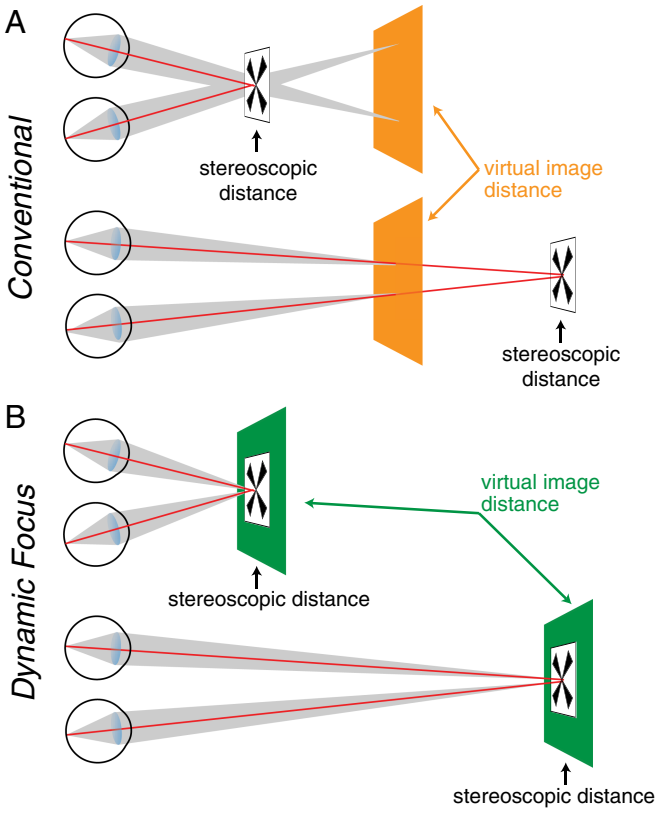

Fig. 2. (*A*) The use of a fixed focus lens in conventional near-eye displays means that the magnified virtual image appears at a constant distance (orange planes). However, by presenting different images to the two eyes, objects can be simulated at arbitrary stereoscopic distances. To experience clear and single vision in VR, the user's eyes have to rotate to verge at the correct stereoscopic distance (red lines), but the eyes must maintain accommodation at the virtual image distance (gray areas). (*B*) In a dynamic focus display, the virtual image distance (green planes) is constantly updated to match the stereoscopic distance of the target. Thus, the vergence and accommodation distances can be matched.

inaccurate corrections or modest changes in clarity that were not noticeable in the virtual scene (*SI Appendix* has additional discussion). Future work can incorporate the refractive testing directly into the system by also using the focus-tunable lenses to determine the spherical lens power that results in the sharpest perceived image and then, store this information for future sessions.

Driving the Eyes' Natural Accommodative Response Using Dynamic Focus. Even in the absence of an uncorrected refractive error, near-eye displays suffer from the same limitations as any conventional stereoscopic display: they do not accurately simulate changes in optical distance when objects move in depth (Fig. 2*A*). To fixate and fuse stereoscopic targets at different distances, the eyes rotate in opposite directions to place the target on both foveas; this response is called vergence (red lines in Fig. 2*A*). However, to focus the displayed targets sharply on the retinas, the eyes must always accommodate to the virtual display distance (gray lines in Fig. 2*A*). In natural vision, the vergence and accommodation distances are the same, and thus, these two responses are neurally coupled. The discrepancy created by conventional near-eye displays (the vergence–accommodation conflict) can, in principle, be eliminated with an adaptive focus display by producing dynamic focus: constantly updating the virtual distance of a target to match its stereoscopic distance (Fig. 2*B*) (19, 20).

Using the autorefractor integrated in our system (Fig. 1*B*), we examined how the eyes' accommodative responses differ between conventional and dynamic focus conditions and in particular, whether dynamic focus can drive normal accommodation by restoring correct focus cues. Users (n = 64, ages 22–63 y old) viewed a Maltese cross that moved sinusoidally in depth between 0.5 and 4 D at 0.125 Hz (mean = 2.25 D, amplitude = 1.75 D), while the accommodative distance of the eyes was continuously measured. Users who wore glasses were tested as described previously with the NETRA, and their correction was incorporated. In the conventional condition, the virtual image distance was fixed at 1.3 m; in the dynamic condition, the virtual image was matched to the stereoscopic distance of the target. Because of dropped data points from the autorefractor, we were able to analyze 24 trials from the dynamic condition, which we compare with 59 trials for the conventional condition taken from across all test groups.

The results are shown in Fig. 3 *A* and *B*. Despite the fixed accommodative distance in the conventional condition, on average, there was a small accommodative response (orange line in Fig. 3*A*) (mean gain = 0.29) to the stimulus. This response is likely because of the cross-coupling between vergence and accommodative responses (24). However, the dynamic display mode (green line in Fig. 3*B*) elicited a significantly greater accommodative gain (mean = 0.77; partially paired one-tailed Wilcoxon tests, $p < 0.001$), which closely resembles natural viewing conditions (25). These results show that it is indeed possible to drive natural accommodation in VR with a dynamic focus display (*SI Appendix* has supporting analysis).

The ability to accommodate degrades with age (i.e., presbyopia) (26). Thus, we examined how the age of our users affected their response gain. For both conditions, accommodative gain was significantly negatively correlated with age (Fig. 3*C*) (conventional $r = -0.34$, dynamic $r = -0.73$, $ps < 0.01$). This correlation is illustrated further in Fig. 3*C*, *Inset*, in which average gains are shown for users grouped by age ($\leq$45 and >45 y old). Although the gains are much greater for the dynamic condition than conventional among the younger age group, the older group had similar gains for the two conditions. From these results, we predicted that accurate focus cues in near-eye displays would mostly benefit younger users and in fact, may be detrimental to the visual perception of older users in VR. We examine this question below.

Optimizing Optics for Younger and Older Users. A substantial amount of research supports the idea that mitigating the vergence–accommodation conflict in stereoscopic displays improves both perception and comfort, and this observation has been a major motivation for the development of displays that support multiple focus distances (3, 5, 7, 12–15, 27). However, the fact that accommodative gain universally deteriorates with age suggests that the effects of the vergence–accommodation conflict may differ for people of different ages (28–30) and even that multifocus or dynamic display modes may be undesirable for older users. Because presbyopes do not accommodate to a wide range of distances, these individuals essentially always have this conflict in their day to day lives. Additionally, presbyopes cannot focus to near distances, and therefore, using dynamic focus to place the virtual image of the display nearby would likely decrease image quality. To test this hypothesis, we assessed sharpness and fusibility with conventional and dynamic focus in younger ($\leq$45 y old, n = 51) and older (>45 y old, n = 13) users.

For the younger group, sharpness was slightly reduced for closer targets in both conditions. However, for the older group, perceived sharpness was high for all distances in the conventional condition and fell steeply at near distances in the dynamic condition (Fig. 3*D*). A logistic regression using age, condition, and distance showed significant main effects of distance and condition. The distance odds ratio was 0.56 (ci = 0.46–0.69), and the ratio for the dynamic condition was 0.60 (ci = 0.48–0.75; $ps < 0.001$),

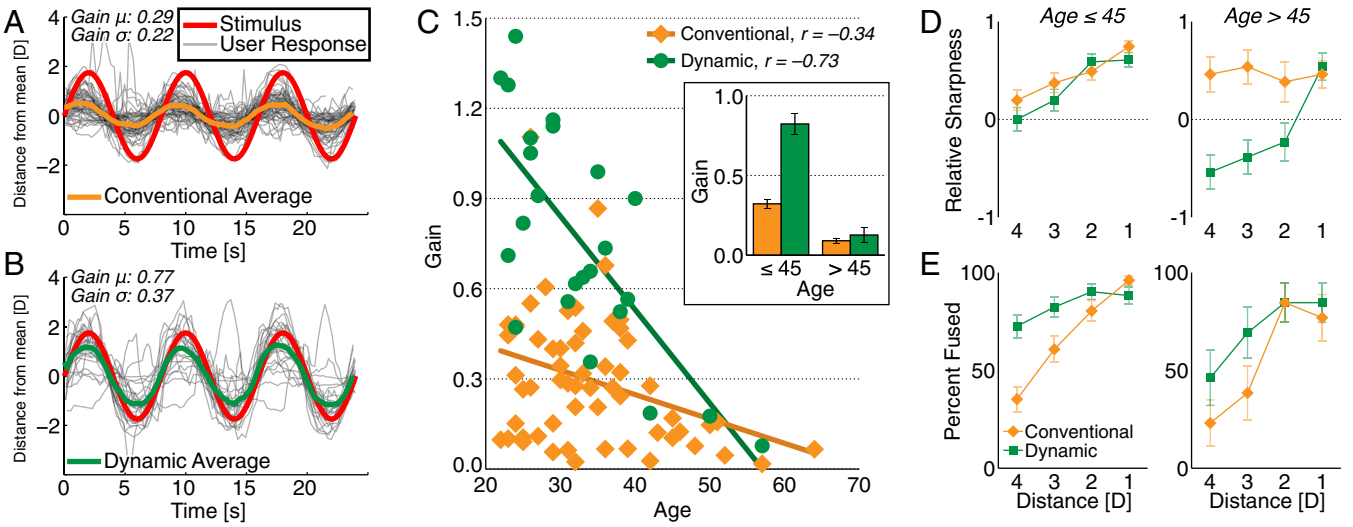

Fig. 3. (*A* and *B*) Accommodative responses were recorded under conventional and dynamic display modes while users watched a target move sinusoidally in depth. The stimulus was shown for 4.5 cycles, and the response gain was calculated as the relative amplitude between the response and stimulus for 3 cycles directly after a 0.5-cycle buffer. The stimulus position (red), each individual response (gray), and the average response (orange indicates conventional focus and green indicates dynamic focus in all panels) are shown with the mean subtracted for each user. Phase is not considered because of manual starts for measurement. (*C*) The accommodative gains plotted against the user's age show a clear downward trend with age and a higher response in the dynamic condition. *Inset* shows means and SEs of the gains for users grouped into younger and older cohorts relative to 45 y old. (*D* and *E*) Average (*D*) sharpness ratings and (*E*) fusibility were recorded for Maltese cross targets at each of four fixed distances: 1–4 D. The *x* axis is reversed to show nearer distances to the left. Error bars indicate SE.

indicating reductions in sharpness at nearer distances. However, the effect of condition was modified by an interaction with age, indicating that sharpness in the older group was reduced significantly more by dynamic mode (odds ratio = 0.70, ci = 0.56–0.87, $p < 0.01$). Indeed, for targets 2 D (50 cm) and closer, older users tended to indicate that the dynamic condition was blurry and that the conventional condition was sharp. The fusibility results for the two age groups were more similar: dynamic focus facilitated fusion at closer distances (Fig. 3*E*). Significant main effects of condition (odds ratio of 1.75, ci = 1.23–2.49) and distance (odds ratio of 0.27, ci = 0.18–0.39) were modified by an interaction (odds ratio of 1.69, ci = 1.27–2.25, all $ps < 0.01$). The interaction indicated that the improvement in fusibility associated with dynamic focus increased at nearer distances. Although dynamic focus provides better fusion for young users, in practice, a more conventional display mode may be preferable for presbyopes. The ideal mode for presbyopes will depend on the relative weight given to sharpness and fusion in determining the quality of a VR experience. In addition, a comfortable focus distance for all images in the conventional condition obviates the need to wear traditional presbyopic correction at all.

We also tested overall preferences while users viewed a target moving in a virtual scene. Interestingly, in both the younger and older groups, only about one-third of the users expressed a preference for the dynamic condition (35% of younger users and 31% of older users). This result was initially surprising given the substantial increase in fusion experienced by younger users in the dynamic mode. One potential explanation is that the target in the dynamic condition may have been modestly less sharp (Fig. 3*D*) and that people strongly prefer sharpness over diplopia. However, two previous studies have also reported overall perceptual and comfort improvements using dynamic focus displays (19, 20). To understand this difference, we considered the fact that our preference test involved a complex virtual scene. Although users were instructed to maintain fixation on the target, if they did look around the scene even momentarily, the dynamic focus (yoked to the target) would induce a potentially disorienting, dynamic vergence–accommodation conflict. That is, unless the dynamic focus is adjusted to the actual distance of fixation, it will likely degrade visual comfort and perception. To address this issue, we built and tested a second system that enabled us to track user gaze and update the virtual distance accordingly.

A Gaze-Contingent Focus Display. Several types of benchtop gaze-contingent display systems—systems that update the displayed content based on where the user fixates in a scene—have been proposed in the literature, including systems that adjust binocular disparity, depth of field rendering, and focus distance (11, 19, 31, 32). Gaze-contingent depth of field displays can simulate the changes in depth of field blur that occur when the eyes accommodate near and far, but they do not actually stimulate accommodation and thus, have not been found to improve perception and comfort (19, 32).

To address the issue of simulating correct accommodative distances in a gaze-contingent manner, we built a second wearable near-eye display system implementing gaze-contingent focus. Our system builds on Samsung's Gear VR platform, but we modify it by adding a stereoscopic eye tracker and a motor that mechanically adjusts the distance between screen and magnifying lenses in real time (Fig. 1*A*, red arrows). To place the virtual image at the appropriate distance in each rendered frame, we use the eye tracker to determine where the user is looking in the VR scene, calculate the distance of that location, and adjust the virtual image accordingly (Fig. 4*A*). This system enabled us to perform comparisons with conventional focus under more naturalistic viewing conditions, in which users could freely look around a VR scene by moving both their head and eyes. Unlike the previous experiments, there was no specific fixation target, and they could move their head to look around the scene. Within each scene, the order of the conditions (conventional, center focus, and gaze-contingent focus) was randomized, and the user was asked to rank them on their perceived image quality.

Based on the insights from our experiments with the benchtop system, we expected users to prefer the gaze-contingent focus condition, particularly when viewing objects at close distances (i.e., 3–4 D). However, if the depth variation in a scene is very

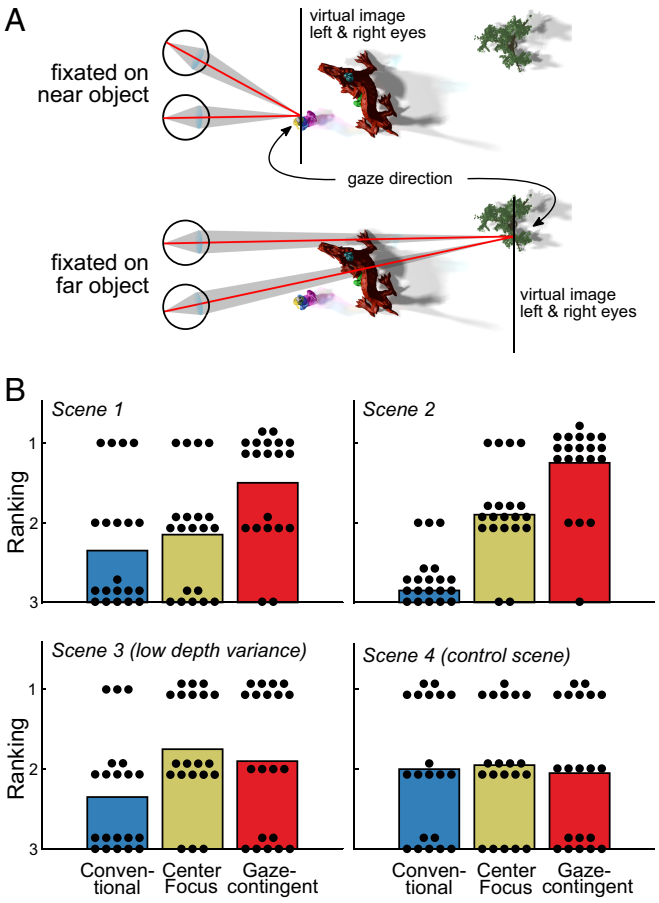

Fig. 4. (*A*) The gaze-contingent focus mode places the virtual image at the same depth as the object in the user's gaze direction. This mode allows the vergence (red lines) and accommodation (gray areas) responses to match at the object depth, which is similar to natural viewing conditions. Note that this illustration does not depict one of the VR scenes used in the experiment (*SI Appendix* shows screenshots of these stimuli). (*B*) User rankings for each condition in four different scenes. Conditions were conventional (the virtual image remains at a fixed optical distance), center focus (the virtual image is adjusted to the scene depth at the center of the view), and gaze-contingent focus. Bar heights indicate average rankings, and black circles show individual responses.

gradual or small, an eye tracker may not be necessary. Instead, the depth of the point in the center of the scene directly in front of the viewer (regardless of whether they are fixating it or not) could be used as a proxy for gaze (center focus). To test these hypotheses, we designed four VR scenes. Scenes 1 and 2 contained large depth variations and nearby objects (up to 4 D). Scene 3 contained objects within a farther depth range (0–2.5 D), and the depth variation was mostly gradual. Finally, scene 4 was a control scene that only contained objects at far distances.

Twenty users (age range = 21–38 y old) ranked three conditions (conventional, center focus, and gaze-contingent focus) for all scenes. We used this age group, because our previous study suggested that younger users would primarily benefit from gaze-contingent focus. As expected, users preferred gaze-contingent focus for scenes with large depth changes and nearby objects (Fig. 4*B*, scenes 1 and 2). Ordinal assessments (Friedman tests) showed a significant effect of condition, and follow-up tests indicated that, for scene 1, gaze-contingent focus was ranked significantly higher than conventional and for scene 2, both center focus and gaze-contingent focus were ranked higher (all $ps < 0.05$, Bonferroni corrected). For the scenes with little to no depth variation (scenes 3 and 4), there was no significant difference between conditions.

Discussion

Near-eye displays pose both a challenge and an opportunity for rethinking visual displays in the spirit of designing a computing platform for users of all ages and abilities. Although the past few years have seen substantial progress toward consumer-grade VR platforms, conventional near-eye displays still pose unique challenges in terms of displaying visual information clearly and comfortably to a wide range of users. The optocomputational displays described here contribute substantially to solving issues of visual quality for users with both normal vision and common refractive errors. Key to these improvements is the idea of an adaptive focus display—a display that can adapt its focus distance in real time to the requirements of a particular user. This adaptive focus can be used to correct near- or far-sightedness and in combination with a mobile eye tracker, can create focus cues that are nearly correct for natural viewing—if they benefit the user. Moving forward, the ability to focus to both far and near distances in VR will be particularly important for telepresence, training, and remote control applications. Similar benefits likely also apply for AR systems with transparent displays that augment the view of the real world. However, AR systems pose additional challenges, because the simulated focus cues for the digital content should also be matched to the physical world.

As an alternative to adaptive focus, two previous studies have examined a low-cost multifocus display solution called monovision (19, 20). In monovision display systems, the left and right eye receive lenses of different powers, enabling one eye to accommodate to near distances and the other eye to accommodate to far distances. We also examined accommodative gain and perceptual responses in monovision (a difference of 1.5 D was introduced between the two eyes, but there was no dynamic focus). We found that monovision reduced sharpness and increased fusion slightly but not significantly, and users had no consistent preference for it. The monovision display also did not drive accommodation significantly more than conventional near-eye displays (*SI Appendix*).

The question of potential negative consequences with long-term use of stereoscopic displays has been raised; however, recent extensive studies have not found evidence of short-term visuomotor impairments or long-term changes in balance or impaired eyesight associated with viewing stereoscopic content (33, 34). In fact, AR and VR near-eye displays have the potential to provide practical assistance to people with existing visual impairments beyond those that are correctable by conventional optics. Near-eye displays designed to provide enhanced views of the world that may increase functionality for people with impaired vision (e.g., contrast enhancement and depth enhancement) have been in development since the 1990s (35, 36). However, proposed solutions have suffered from a variety of limitations, including poor form factor and ergonomics and restricted platform flexibility. The move to near-eye displays as a general purpose computing platform will hopefully open up possibilities for incorporating low-vision enhancements into increasingly user-friendly display systems. Thus, in the future, AR/VR platforms may become accessible and even essential for a wide variety of users.

Materials and Methods

Display Systems. The benchtop prototype uses Topfoison TF60010A Liquid Crystal Displays with a resolution of 2,560 × 1,440 pixels and a screen diagonal of 5.98 in. The optical system for each eye offers a field of view of 34.48° and comprises three Nikon Nikkor 50-mm f/1.4 camera lenses. The focus-tunable lens (Optotune EL-10-30-C) dynamically places the virtual image at any distance between 0 and 5 D and changes its focal length by shape deformation within 15 ms. The highest rms wavefront error exhibited by the lens

placed in a vertical orientation (according to Optotune) is 0.3 λ (measured at 525 nm). No noticeable pupil swim was reported. Two additional camera lenses provide a 1:1 optical relay system that increases the eye relief so as to provide sufficient spacing for a near-IR (NIR)/visible beam splitter (Thorlabs BSW20R). The left one-half of the assembly is mounted on a Zaber T-LSR150A Translation Stage that allows interpupillary distance adjustment. A Grand Seiko WAM-5500 Autorefractor records the accommodation state of the user's right eye at about 4–5 Hz with an accuracy of $\pm$0.25 D through the beam splitter. The wearable prototype is built on top of Samsung's Gear VR platform with a Samsung Galaxy S7 Phone (field of view = 96°, resolution = 1,280 $\times$ 1,440 per eye). A SensoMotoric Instruments (SMI) Mobile ET-HMD Eye Tracker is integrated in the Gear VR. This binocular eye tracker operates at 60 Hz over the full field of view. The typical accuracy of the gaze tracker is listed as $< 0.5°$. We mount an NEMA 17 Stepper Motor (Phidgets 3303) on the SMI Mobile ET-HMD Eye Tracker and couple it to the focus adjustment mechanism of the Gear VR, which mechanically changes the distance between phone and internal lenses. The overall system latency is approximately 280 ms for a sweep from 4 to 0 D (optical infinity). For reference, a typical response time for human accommodation is around 300–400 ms (discussion is in *SI Appendix*) (37).

Experiments. Informed consent was obtained from all study participants, and the procedures were approved by the Stanford University Institutional Review Board. Details are in *SI Appendix*.

Data Availability. Dataset S1 includes the raw data from both studies.

ACKNOWLEDGMENTS. We thank Joyce Farrell, Max Kinateder, Anthony Norcia, Bas Rokers, and Brian Wandell for helpful comments on a previous draft of the manuscript. N.P. was supported by an National Science Foundation (NSF) Graduate Research Fellowships Program. E.A.C. was supported by Microsoft and Samsung. G.W. was supported by a Terman Faculty Fellowship, an Okawa Research Grant, an NSF Faculty Early Career Development (CAREER) Award, Intel, Huawei, Samsung, Google, and Meta. Research funders played no role in the study execution, interpretation of data, or writing of the paper.

1. Wheatstone C (1838) Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos Trans R Soc Lond* 128:371–394.
2. Sutherland IE (1968) A head-mounted three dimensional display. Proceedings of Fall Joint Computer Conference (ACM, New York), pp 757–764.
3. Kooi FL, Toet A (2004) Visual comfort of binocular and 3D displays. *Displays* 25:99–108.
4. Lambooij M, Fortuin M, Heynderickx I, IJsselsteijn W (2009) Visual discomfort and visual fatigue of stereoscopic displays: A review. *J Imaging Sci Technol* 53(3):1–14.
5. Shibata T, Kim J, Hoffman DM, Banks MS (2011) The zone of comfort: Predicting visual discomfort with stereo displays. *J Vis* 11(8):11.
6. Cutting JE, Vishton PM (1995) Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. *Perception of Space and Motion*, eds Epstein W, Rogers S (Academic Press, San Diego), pp 69–117.
7. Hoffman DM, Girshick AR, Akeley K, Banks MS (2008) Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J Vis* 8(3):1–30.
8. Vitale S, Ellwein L, Cotch M, Ferris F, Sperduto R (2008) Prevalence of refractive error in the United States, 1999-2004. *Arch Ophthalmol* 126(8):1111–1119.
9. Duane A (1912) Normal values of accommodation at all ages. *J Am Med Assoc* LIX(12):1010–1013.
10. Traub AC (1967) Stereoscopic display using rapid varifocal mirror oscillations. *Appl Opt* 6(6):1085–1087.
11. Shiwa S, Omura K, Kishino F (1996) Proposal for a 3-D display with accommodative compensation: 3DDAC. *J Soc Inf Disp* 4(4):255–261.
12. Rolland J, Krueger M, Goon A (2000) Multifocal planes head-mounted displays. *Appl Opt* 39(19):3209–3215.
13. Akeley K, Watt S, Girshick A, Banks M (2004) A stereo display prototype with multiple focal distances. *ACM Trans Graph* 23(3):804–813.
14. Liu S, Cheng D, Hua H (2008) An optical see-through head mounted display with addressable focal planes. Proceedings of ISMAR (IEEE Computer Society, Washington, DC), pp 33–42.
15. Love GD, et al. (2009) High-speed switchable lens enables the development of a volumetric stereoscopic display. *Opt Express* 17(18):15716–15725.
16. Lanman D, Luebke D (2013) Near-eye light field displays. *ACM Trans Graph* 32(6):1–10.
17. Huang FC, Chen K, Wetzstein G (2015) The light field stereoscope: Immersive computer graphics via factored near-eye light field display with focus cues. *ACM Trans Graph* 34(4) 60:1–12.
18. Banks MS, Hoffman DM, Kim J, Wetzstein G (2016) 3D displays. *Annu Rev Vis Sci* 2(1):397–435.
19. Konrad R, Cooper EA, Wetzstein G (2016) Novel optical configurations for virtual reality: Evaluating user preference and performance with focus-tunable and monovision near-eye displays. ACM CHI Conference on Human Factors in Computing System (ACM, New York), pp 1211–1220.
20. Johnson PV, et al. (2016) Dynamic lens and monovision 3D displays to improve viewer comfort. *Opt Express* 24:11808–11827.
21. Llull P, et al. (2015) Design and optimization of a near-eye multifocal display system for augmented reality. *OSA Imaging and Applied Optics 2015* (Optical Society of America, Washington, DC), p JTH3A.5.
22. World Health Organization (2014) *Visual Impairment and Blindness*. Available at www.who.int/mediacentre/factsheets/fs282/en/. Accessed September 29, 2016.
23. Pamplona VF, Mohan A, Oliveira MM, Raskar R (2010) Netra: Interactive display for estimating refractive errors and focal range. *ACM Trans Graph* 29(4):77:1–77:8.
24. Fincham EF, Walton J (1957) The reciprocal actions of accommodation and convergence. *J Physiol* 137(3):488–508.
25. Charman WN, Heron G (2000) On the linearity of accommodation dynamics. *Vision Res* 40(15):2057–2066.
26. Heron G, Charman WN (2004) Accommodation as a function of age and the linearity of the response dynamics. *Vision Res* 44(27):3119–3130.
27. Schowengerdt BT, Seibel EJ (2006) True 3-D scanned voxel displays using single or multiple light sources. *J Soc Inf Disp* 14(2):135–143.
28. Watt S, Ryan L (2015) Age-related changes in accommodation predict perceptual tolerance to vergence-accommodation conflicts in stereo displays (abs.). *J Vis* 15:267.
29. Yang SN, et al. (2012) Stereoscopic viewing and reported perceived immersion and symptoms. *Optom Vis Sci* 89(7):1068–1080.
30. Read JC, Bohr I (2014) User experience while viewing stereoscopic 3D television. *Ergonomics* 57(8):1140–1153.
31. Peli E, Hedges TR, Tang J, Landmann D (2012) 53.2: A binocular stereoscopic display system with coupled convergence and accommodation demands. *SID Dig* 32(1):1296–1299.
32. Mauderer M, Conte S, Nacenta MA, Vishwanath D (2014) Depth perception with gaze-contingent depth of field. ACM CHI Conference on Human Factors in Computing System (ACM, New York), pp 217–226.
33. Read JC, et al. (2015) Viewing 3D TV over two months produces no discernible effects on balance, coordination or eyesight. *Ergonomics* 59(8):1073–1088.
34. Read JC, et al. (2015) Balance and coordination after viewing stereoscopic 3D television. *R Soc Open Sci* 2(7):140522.
35. Massof RW, Rickman DL (1992) Obstacles encountered in the development of the low vision enhancement system. *Optom Vis Sci* 69(1):32–41.
36. van Rheede JJ, et al. (2015) Improving mobility performance in low vision with a distance-based representation of the visual scene. *Invest Ophthalmol Vis Sci* 56(8):4802.
37. Campbell FW, Westheimer G (1960) Dynamics of accommodation responses of the human eye. *J Physiol* 151:285–295.

Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays

Padmanaban et al.

Supporting Information (SI)

Review of Related Display Technologies.

A comprehensive overview of 3D displays in general and related perceptual issues can be found in the recent survey by Banks et al. (1). In the following, we briefly review display technologies that are aimed at enhancing 3D information via focus cues.

Retinal Blur and Disparity Rendering Software-only approaches for attempting to provide focus cues have been explored in the past. For example, dynamic depth of field rendering is a software-only approach that renders the fixated object sharply while blurring other objects according to their relative distance in the virtual environment (2–4). There are no optical changes to the display in this mode—the distance of the virtual image remains fixed as in the conventional mode, only the displayed images change. In all studies (2–4), this technique has not been found to be perceptually beneficial, most likely because human accommodation may be driven by the accommodation-dependent blur gradient. Specifically, in natural conditions, changes of perceived retinal blur with respect to changes in the accommodative state of the crystalline lens would drive a user's accommodation to the distance of the fixated object. The retinal blur gradient in a digital display drives accommodation to its virtual image, no matter how the displayed content is altered.

Gaze-contingent disparity rendering (5) is another software-only approach that is effective in retaining the relative disparity cues between rendered objects by shifting the the entire 3D scene in depth such that the fixated object appears at the zero-disparity plane (i.e., the virtual image distance).

Limitations: Neither of these software-only approaches provide natural focus cues nor do they make near-eye displays accessible for users with refractive errors.

Volumetric and Multi-plane Displays Three-dimensional volumetric and multi-plane displays represent the most common approach to focus-supporting near-eye displays. Instead of using 2D display primitives at some fixed or adaptive distance to the eye, volumetric displays either mechanically or optically scan out the 3D space of possible light emitting display primitives in front of each eye (6).

Multi-plane displays approximate this volume using a few virtual planes that are generated by beam splitters (7, 8) or time-mulitplexed focus-tunable optics (9–15). According to Campbell (16), the depth of focus, the change in focal distance that is just noticeable, under normal viewing conditions is roughly 0.3 D, which could be used to determine the spacing between the presentation planes. Interestingly, MacKenzie et al. (17, 18) recently used multi-focal displays to determine that a spacing of 0.9 D between discrete focal planes is sufficient to create accurate and natural accommodation responses in a multi-plane display. Our study does not evaluate multi-plane displays—both of our prototypes allow for continuous focus adjustment in the adaptive display mode. However, our prototypes could easily be adapted to offer a limited set of directly addressable focal planes. In this case, the insights of MacKenzie et al.'s study would also apply and can be used as a guideline.

Limitations: Implementations with beam splitters make small device form factor for wearable near-eye displays difficult, temporal multiplexing can introduce perceived flicker and requires display refresh rates beyond those offered by current-generation microdisplays. Without significant improvements in refresh rates of displays commonly used for VR/AR applications, volumetric and multi-plane displays do not seem practical.

Light Field and Holographic Displays Other displays, such as light field and holographic displays, aim to synthesize the full 4D light field in front of each eye. Conceptually, this approach allows for parallax over the entire eye box to be accurately reproduced, including monocular occlusions, specular highlights, and other effects that cannot be reproduced by volumetric displays.

Limitations: Current-generation light field displays provide limited resolution (19–21) whereas holographic displays suffer from speckle and have extreme requirements on pixel sizes that are not afforded by near-eye displays also providing a large field of view.

Virtual Retinal Displays Direct stimulation of the retina, for example with scanned laser projectors, has been proposed. By focusing the laser through a small region of the user's pupil, the retinal blur is basically non-existent and can be controlled in software. This display mode renders the accommodative system into an open loop condition.

Limitations: Retinal displays in open loop do not drive human accommodation and they suffer from a very small eye box. Users of such systems often perceive visual artifacts, caused by shadows of small particles or aberrations in their own eyes. These and other limitations have prevented widespread use of virtual retinal displays.

Optical See-through Augmented Reality (AR) Displays AR displays are closely related to but different from virtual reality (VR) displays. In VR, the user's visual field is only stimulated by digital content created by a near-eye display. AR displays provide similar stimuli, but optically superimpose these on the physical stimuli of the real world. Most commonly, AR displays use optical combiners similar to those used in Pepper's Ghost illusion to combine digital and physical stimuli.

The goal of focus-supporting displays is somewhat different for VR and AR displays. In VR, the goal for non-presbyopes is to drive accommodation so as to provide natural viewing conditions and mitigate the vergence–accommodation conflict and its symptoms. In AR applications, there are two goals. The first is the same as for VR: one would like to be able to drive accommodative user responses with the digital content. Second, one would also like to ensure consistent focus cues between physical and digital content. The first goal for AR is adequately addressed with our VR-centric study. Evaluating

consistency of focus cues in AR is an interesting avenue of future research.

Benchtop Focus-tunable Display System.

Hardware An additional image of this prototype is shown in Fig. S1. The optical system for each eye comprises three Nikon Nikkor 50-mm f/1.4 camera lenses and a focus-tunable lens (Optotune EL-10-30-C with 10 mm diameter and a focal range of 5 to 10 diopters (D)). The Nikon lens closest to the screen is mounted at its focal distance to create a virtual image at optical infinity. The focus-tunable lens abuts the Nikon lens closest to the screen, and is offset with a –10 D lens. Without current applied, the focus-tunable lens places the virtual image at 5 D (0.2 m), but with increasing current the curvature of the liquid lens is increased, thereby placing the virtual image at a farther distance from the observer. Hence, the other two camera lenses provide a 1:1 optical relay system that increases the eye relief to about 4–5 cm. Without this relay, the eyeball of the observer would have to be placed right on the focus-tunable lens, which is mechanically difficult and would also not be comfortable. The eye relief also provides space for a visible/near-infrared beam splitter (Thorlabs BSW20R) in front of the eyes, so as to allow the autorefractor to function.

A Grand Seiko WAM-5500 autorefractor is integrated into the near-eye display system. The autorefractor uses built-in near infrared illumination (NIR) and a NIR camera to determine the user's accommodative state. The illumination pattern is close to invisible to the user. Accommodation measures are directly transmitted to the computer that controls the visual stimulus. The accuracy of the autorefractor is verified using a Heine Ophthalmoscope Trainer model eye (C-000.33.010).

Software All software driving the prototype is implemented in C++. The OpenGL application programming interface is used for 3D rendering.

Calibration. The prototype is calibrated in several steps to ensure that the images presented to the users exhibited no vertical parallax and are displayed with the right optical parameters. At each of the following steps, a bubble level is used to ensure that all parts are level. First, a digital single-lens reflex camera (DSLR) is used to verify that the screen is indeed placed at infinity by the lens mounted closest to the screen. Then, once the left and right assemblies are completed, a set of 4 calibration targets are used to ensure that the VIS/NIR mirror is placed at a 45° angle in each optical path. Two of the targets are placed right after the screen, and two in the perpendicular direction relative to the screen after the VIS/NIR beam splitters. The beam splitter positions and rotations are adjusted such that the full field of view from the lenses is reflected, and all 4 calibration targets are aligned in the optical axis. With just the calibration targets at the screen in place, the position of the 600 px square viewport on the LCDs is adjusted such that the onscreen targets are centered through the calibration targets when they are displayed at infinity. We then verify that the viewport is roughly equal in size to the circular field of view.

To set the height of the two assemblies, we first adjust the right assembly's height until a representative user's eye is vertically centered in the autorefractor's display as they look through the beam splitters. The left assembly is then adjusted to the same height (verified by connecting a cage rod across the two assemblies and a bubble level). Here the cage rod also serves to verify that the two assemblies are parallel to each other. While displaying the Maltese cross target on both screens, we use a camera and adjust the forward tilt of the beam splitters until the two targets are at the same vertical position. Any small amount (a pixel or two) of vertical parallax remaining is fixed by moving the viewports on screen. Lack of vertical parallax is then verified with a human user.

Finally, we calibrate the power of the Optotune lenses and the position of the Zaber stage so that the lens power and interpupillary distance (IPD) can be programmatically controlled. For the Optotune lenses, we place a DLSR at the user's viewing position, set its focal distance at each of the marked distances on its lens, and for each distance measure the focal powers output by the driver. The offset to apply in software is calculated using a best fit linear regression between Optotune driver and DSLR focal powers in diopters. The Zaber stage is calibrated by setting its maximum position where the two assemblies are at their nearest possible distance, then recording the distance between the optical axes (identified as above using 4 calibration targets).

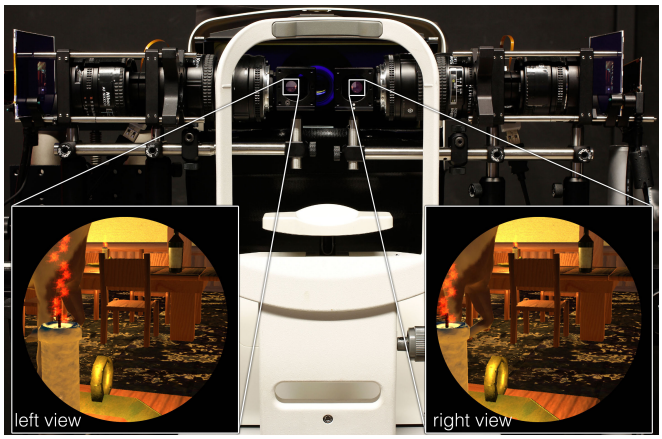

Fig. S1. Prototype stereoscopic near-eye display with focus-tunable lenses and adjustable interpupillary distance via a translation stage. The systems includes an autorefractor that is capable of recording the accommodative state of the user's right eye continuously at 4–5 Hz. Insets show example stereoscopic views.

Gaze-contingent Near-eye Display System.

Hardware Fig. S2 shows a schematic view of this hardware setup. The system provides a 96° field of view, a significant improvement over the benchtop variant. The NEMA 17 stepper motor mounted on the Samsung Gear VR is coupled via a 2" pulley and an O-ring to the focus adjustment mechanism, which mechanically changes the distance between phone and internal lenses. The motor offers a 1.8° step angle, with a 0.1125° micro-step angle, and 4.24 kg·cm torque by drawing up to 2.5 A of current. The driver board (Phidget 1067 PhidgetStepper Bipolar HC) is controlled by a Raspberry Pi 3 Model B that is connected to the Gear VR phone via bluetooth. Overall system latency, including rendering, data transmission, and motor adjustments, are approx. 280 ms for a sweep from 4 D (25 cm) to 0 D (optical infinity) and 160 ms for a sweep from from 3 D (33 cm) to 1 D (1 m). The system's latency can be noticeable for a user who knows what to look for, even though the response time of the human accommodative re-

Fig. S2. A conventional near-eye display (Samsung Gear VR) is augmented by a gaze tracker and a motor that is capable of adjusting the physical distance between screen and lenses.

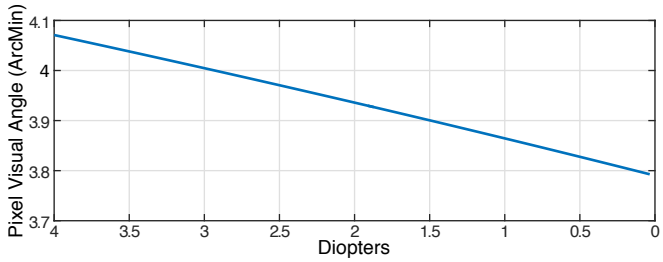

Fig. S3. The visual angle of one pixel of a Samsung Galaxy S7 as perceived by a user with 2 cm eye relief, as a function of the distance to the virtual image.

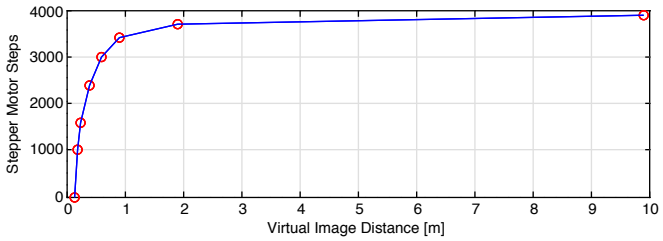

Fig. S4. We measure the mapping between stepper motor micro-steps, each corresponding to a 0.1125° rotation, to distances of the virtual image (red circles). A high-resolution lookup table is then computed by linearly interpolating between the measurements (blue line).

sponse is also on the order of hundreds of milliseconds (22). Nonetheless, the user study with this system indicates that the system latency does not remove the benefits of adaptive modes for user preferences. Indeed, reducing the latency to a point at which it was unobtrusive was one of the engineering challenges posed by designing this system.

Due to the user's eye being spaced a small distance away from the lens, some amount of magnification change necessarily occurs when changing the distance to the virtual image. To determine how large this change is, we calculated the visual angle subtended by one pixel as a function of the virtual image distance (Fig. S3). Over a 4 D range, the visual angle of one pixel changes by less than 0.3 arcminutes. These changes can be compensated for in software, at the cost of potentially introducing aliasing artifacts, or optically, at the cost of introducing a tunable lens into the system. However, given how small the magnification change is in the current system, we decided to evaluate it without compensation.

Software All applications are written in Unity version 5.4.1. SMI provides a Unity plugin for their eye tracker, which allows for control of all components from a single application. Unity is connected to the Raspberry Pi via bluetooth, but only a single metric distance value needs to be transmitted per rendered frame which is easily possible at the available bandwidth. The Raspberry Pi uses Python code and a Python-based software development kit provided by Phidgets to drive the stepper motor to the desired distance.

Calibration As a one-time pre-processing step, we calibrate the mapping between stepper motor settings and metric distances of the virtual image. This calibration is done by focusing a DSLR camera at several distances through the Gear VR optics and adjusting the motor such that the virtual image appears at each of these distances. Linearly interpolating between the sparse set of calibrated distances, a high-resolution

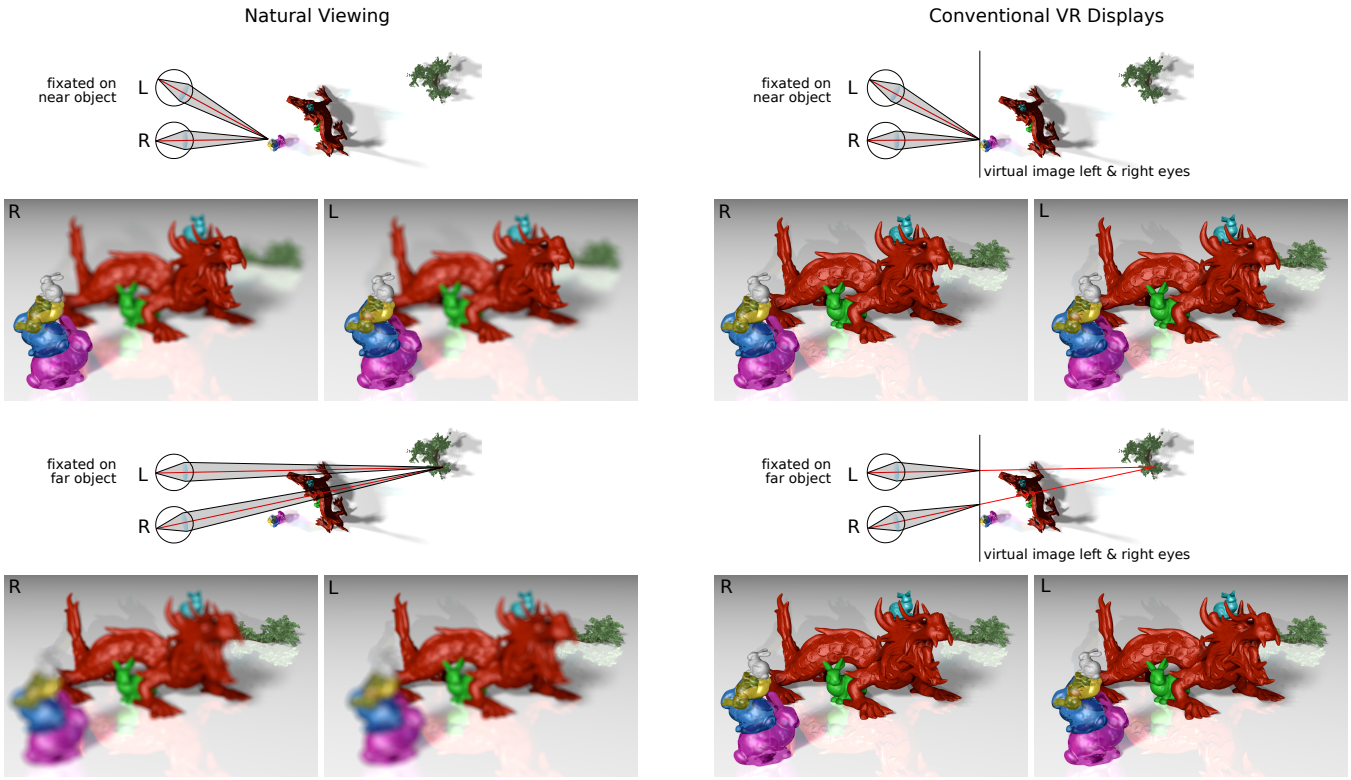

Fig. S5. Each panel includes a top down view of a fixation point and a cross-fusable stereo pair. In natural viewing conditions (left), the eyes verge (red lines) and accommodate (gray areas) at the depth of the fixated object. Current-generation near-eye displays (right) create a magnified virtual image to which the user accommodates. For virtual objects displayed at distances away from that virtual image, vergence and accommodation cues are in an unnatural state known as the vergence–accommodation conflict. Users whose accommodation range does not include the virtual image distance, e.g. due to myopia, hyperopia, or presbyopia, cannot see a sharp image in conventional VR displays at all.

lookup table is computed that maps virtual image distances to specific motor settings (Fig. S4). This lookup table is stored on the Raspberry Pi and used to drive the virtual image to a specific metric distance, as requested by the Unity application.

The eye tracker is calibrated using a per-user, per-use procedure that is partly provided by the manufacturer. In this procedure, the user is presented with several targets appearing at different locations within their visual field and asked to fixate on these targets. For the gaze-contingent focus display, we repeat this procedure three times and adjust the virtual image distance to either 0.67, 3, or 4 D for each of these runs. During runtime, we dynamically select the calibration setting that is closest to the set of pre-calibrated virtual image distances. This calibration makes gaze tracking more robust than calibrating only for a single virtual image distance when dynamically changing the virtual image distance.

Summary of display modes. The following summarizes the different display modes that were evaluated in this study.

Conventional mode is the display mode employed by current-generation near-eye displays. Simple magnifying lenses enlarge the image of a microdisplay and create a virtual image at some fixed distance to the viewer. Fig. S5 illustrates how conventional stereoscopic displays differ from natural viewing. Top-down views illustrate the vergence and accommodative distances when looking at objects at different distances in the physical world (left) and in a conventional stereoscopic display (right). Cross-fusable stereo pairs simulate the visual cues (binocular disparity and retinal blur) available during each fixation. In natural viewing, the accommodation state of the eye also creates the perceived retinal blur or depth of field cue. This depth of field is different for near and far fixation distances (Fig. S5, left). For conventional VR, the accommodative state does not change and is fixed on the virtual image. Hence, all displayed objects are equally sharp (Fig. S5, right), unless the depth-dependent blur is rendered into the image.

Corrected mode is the simplest adaptive display mode. The virtual image is kept at a fixed distance, but this distance is corrected to account for the user's refractive error, either hyperopia and myopia.

Dynamic mode is a software/hardware approach that adjusts the virtual image distance, either by changing the focal length of the lenses or the distance between the microdisplay and the lenses. No eye tracking is used in this mode, so the virtual image distance is either linked to some specific rendered object in the scene or to the principle direction of head orientation, as for example determined by an inertial measurement unit in a head mounted display. When we evaluate the latter approach in the primary text, we call it *center focus* to differentiate it from the former.

Gaze-contingent focus mode is a software/hardware approach that adjusts the virtual image distance based on the depth at which the viewer is fixated, either by changing the focal length of the lenses or the distance between the microdisplay and the lenses. With information from an eye tracker, this mode allows for the virtual image distance to be adjusted

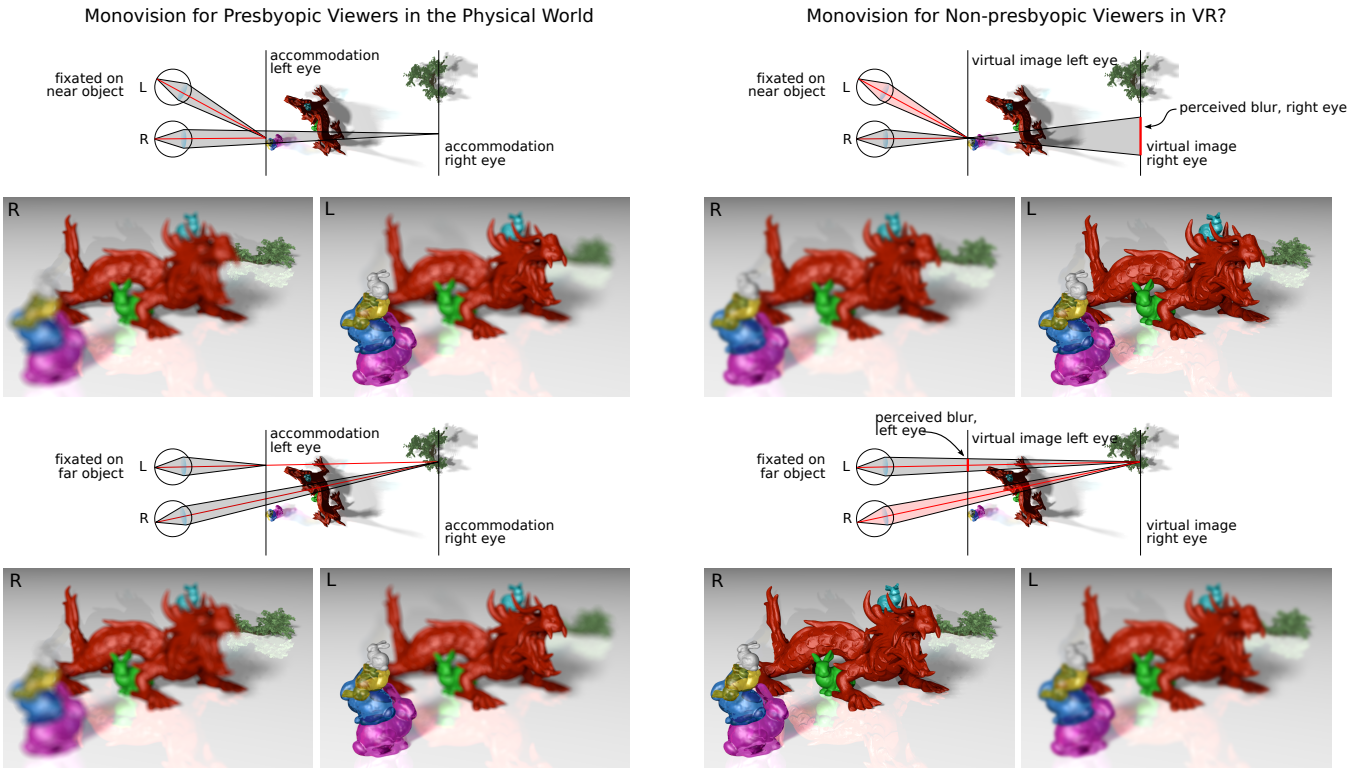

Fig. S6. Monovision is a common vision correction technique for presbyopic persons in the physical world. For this purpose, the dominant eye is usually corrected for far vision and the other eye for near vision, such that the presbyopic user is provided with sufficient visual acuity for different viewing conditions (left). It was recently hypothesized (4, 23, 24) that monovision integrated into near-eye displays may be able to allow non-presbyopic viewers to accommodate at two discrete distances (right). Assuming that the accommodation state of both eyes is coupled, for the VR monovision to work as intended the eye that drives accommodation for both eyes (red shaded area on the right) would have to switch depending on the vergence state of the eyes. Similar to monovision for presbyopic viewers, one eye would always see a blurry image whereas the other would see a sharp image; the order would switch when fixating on near or far objects. However, applied to these two applications, monovision has a very different goal and may only be useful for two disparate groups of people.

to either the distance where the viewer is verged (requires vergence tracking) or at the depth corresponding to their gaze direction (requires gaze tracking). In our study, we implement this mode with gaze tracking.

Monovision refers to a common treatment for presbyopia, a condition that often occurs with age in which people lose the ability to focus their eyes on nearby objects. To improve visual clarity, monovision places two lenses with different prescription values in front of each eye such that one eye dominates for distance vision and the other for near vision (Fig. S6, left). Monovision was recently proposed for VR/AR applications (4, 23, 24) and is evaluated in more detail in the following sections (Fig. S6, right).

Study design and details.

Users For the main study, data were collected from volunteers who participated as part of a public demonstration at the SIGGRAPH 2016 conference. The gender make-up of the users was 127 males, 25 females, and 1 not stated. All users filled out a questionnaire indicating their age, gender, familiarity with virtual reality, typical optical correction, and any other issues affecting their vision. Of the original study volunteers, eleven users indicated having ocular issues beyond the common target conditions and were thus excluded from the study. Another five users were excluded due to significantly incomplete questionnaires. Users were randomly assigned to one of three groups: corrected versus uncorrected, conventional versus dynamic, or conventional versus monovision. Users who did not need optical correction or were wearing contact lenses were not assigned to the first group. Two more users were excluded due to incorrect assignment to conditions on the part of the experimenter.

A follow-up study was conducted at Stanford University. In the followup study, users were asked to perform a stereovision test to verify stereoacuity of at least 40 arcsec at 16 in (Randot test). They were also asked to perform a near point test in which we measured the nearest distance at which they could clearly and comfortably view a document to verify that they had a typical near point of 25 cm (4 D). They wore their corrective eyewear during the test. Three users were excluded prior to data collection due to not meeting the threshold in the random dot or near point tests. The remaining users were made up of 15 males and 5 females (age range 21–38 y old). All remaining users had corrected vision and viewed all three conditions (conventional, center focus, gaze-contingent focus). In all cases, the order of presentation of conditions was randomized per user and per task. One user from the follow up study also participated in the tasks for the main study.

Ocular Measurements In the main study, users had their near- and far-correction measured using the EyeNetra NETRA. The NETRA is a portable, cellphone-based autorefractor that measures spherical and cylindrical correction as well as interpupillary distance (IPD) (25).

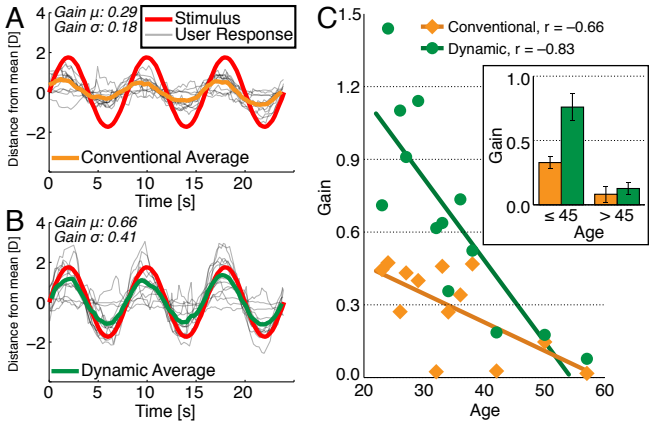

Fig. S7. Accommodative responses were recorded under conventional and dynamic display modes while users watched a target move sinusoidally in depth. Data are plotted in the same manner as Fig. 3, but only users with valid paired data for dynamic and conventional modes are shown. (*A* and *B*) The stimulus position (red), each individual response (gray), and the average response (orange indicates conventional and green indicates dynamic focus in all panels), are shown with the mean distance subtracted. (*C*) These gains plotted against the user's age show a clear downward trend with age, and a higher response in dynamic. Inset shows mean and standard error of the gains for users grouped into younger and older cohorts relative to forty-five years old.

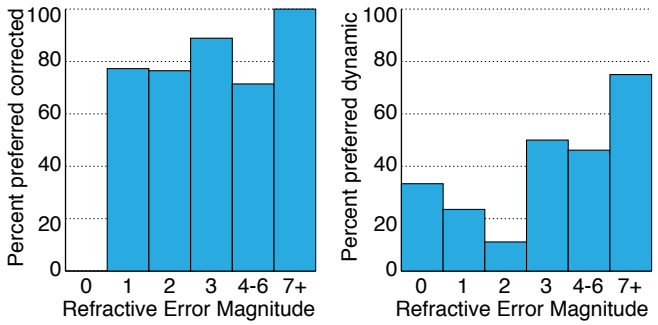

Fig. S8. The reported preference data for users is shown for users that experienced the corrected and uncorrected conditions (left), and those that experienced the conventional and dynamic conditions (right), as a function of the the magnitude of their highest refractive error (between eyes). Each bar indicates the percent of users that preferred the condition hypothesized to be more preferable (i.e., corrected and dynamic on the left and right, respectively). Note that the histogram bins for the corrected/uncorrected preferences are chosen to start at 0.5 D because users with lower measured errors were considered emmetropic and not placed in this condition.

For those wearing contacts, only the IPD was used. Users whose spherical equivalent (spherical + 0.5 × cylindrical) measurement was less than ±0.5 D were considered emmetropic and were not assigned to the corrected versus uncorrected condition. Of those not considered emmetropic, 114 were myopic, 11 hyperopic, and 2 were anisometropic such that one eye was myopic and the other hyperopic. In the study comparing conventional and dynamic conditions, none of the users in the older age group indicated that they were wearing multifocal contact lenses.

Accommodation measures Users were presented with a Maltese cross target, which they were asked to follow with their eyes. The target was scaled to be 6.2 cm in height. When the users indicated readiness, the experimenter started the cycle for one condition, then repeated this process again for the other condition (randomized order), while simultaneously and manually starting measurement with the autorefractor each time.

Due to frequent erroneous data points from the autorefractor that manifested as large values or as time stamps with missing data points, several users' measurements were excluded from analysis, leaving data from 83 users. First, erroneous data points were identified as those that spiked to more than twice the stimulus amplitude from the median of the user's response for that mode. These were verified to not have removed any points from users that may have had a large voluntary accommodation. Any measurements in which 15% or more data points were removed as erroneous were excluded entirely. The remaining measurements were viewed without display mode labels, and any measurements containing spikes characteristic of the autorefractor's errors were excluded in their entirety as well. Finally, measurements with fewer than 6 data points were automatically excluded.

A gain value was calculated for each of the remaining measurements using Fourier analysis. First, we extracted the first 3 cycles of measurement data, following the first 0.5 cycles which were a buffer region, then subtracted the mean for these 3 cycles. We applied a Hamming window and projected the data points onto an 0.125 Hz exponential sinusoid at the time points provided by the autorefractor (which has non-uniform sampling times). The magnitude of the projection was then scaled according to the stimulus amplitude to give the gain. Due to excluded measurements, we use a partially paired analysis to pool the paired and unpaired data between the conventional and dynamic conditions (26), calculated using Wilcoxon signed-rank and rank-sum tests, respectively. The weights for pooling were calculated using the size of the smaller dataset.

Because some data were excluded, the subsequent gain analysis comparing conventional and dynamic conditions included both paired and unpaired measures. To confirm that the differences we see between these two conditions are not attributable to between-subjects differences, in Fig. S7, we consider the data for only the subset of users for whom both conditions were analyzed. Data are plotted in the same manner as Fig. 3 in the main paper, and the patterns in these data are the same as for the full analysis. In fact, within these users, the accommodative gain elicited by the dynamic condition was always greater than that elicited by the conventional condition, further supporting the finding that the dynamic mode elicits higher accommodative responses.

Although not reported in the main paper, accommodative gains were also measured for users in the corrected vs. uncorrected study. The average gain for users in the uncorrected was 0.27, as compared to the gain in the corrected/conventional mode of 0.29 as stated in the main paper. This difference was not statistically significant.

Sharpness and fusion tests To measure sharpness and fusion, users were instructed to focus on the Maltese cross target displayed at a fixed distance of 1, 2, 3, or 4 D and try to fuse it into a single image (it was on-screen for 4 s, with a 4 s delay before appearing). The target was always displayed such that it subtended 3.45° of the visual field, and all trial types (condition and distance) were randomized. A response screen appearing after each trial prompted the user to respond as to whether the target was fused, and to rate it as blurry, medium, or sharp.

Data were analyzed using mixed effects logistic regressions, with user modeled as a random effect. To evaluate the correction condition, we used a 2×4 model (condition $\times$ distance). To evaluate the dynamic condition, we used a $2 \times 2 \times 4$ model (condition $\times$ age group $\times$ distance).

Preference test The users were presented their assigned two conditions in a random order, the first with a target labeled "A," and the second with a target labeled "B." They were asked to freely toggle between the two modes as they fixated on the target, then choose the one that they found more comfortable to observe. The targets were 6.2 cm tall and moving at 0.125 Hz sinusoidally in depth. The first 84 users were shown targets that moved between 25 and 75 cm (4 and 1.33 D), while the later users saw the targets move between 33.3 and 83.3 cm (3 and 1.2 D).

The results were analyzed using a binomial test. The results were first analyzed separately for the first 84 and the remaining users. For the uncorrected vs corrected condition preference, 81% of the first cohort preferred corrected and 79% of the second cohort preferred corrected (both $ps < 0.001$). Since the proportions were similar, we pooled all data before analysis. For the dynamic vs conventional preferences, within group differences were also highly similar so all users were pooled.

We also examined the potential effect of refractive error on preferences. For instance, because the conventional condition places the virtual image at a distance of 1.3 m, or 0.77 D, a user with myopia of –1 D would effectively see a clear image even without refractive correction. For the two studies, Fig. S8 shows the percent of users that preferred the corrected over the uncorrected condition (left) and the percent of users that preferred the dynamic over the conventional condition (right). Although not conclusive, this analysis suggests that users with only a small amount of correction were slightly closer to chance in their preferences for corrected versus uncorrected, whereas users with large refractive corrections were more likely to prefer the corrected condition. Interestingly, the analysis also suggests that users with larger refractive errors may be more likely to prefer the dynamic condition. However, the reason for this difference is unclear.

Gaze-contingent preference test In the follow-up study using the wearable prototype, users were presented with 4 different VR scenes and asked to rank the display modes in each. This study focused on image quality rather than comfort based on the hypothesis that the study duration was likely too short to affect comfort. The user was again free to toggle through the conditions repeatedly before ranking them. They were not given a specific fixation target and could explore the scene visually and move their head. In order to reduce the user's ability to differentiate the conventional mode based on motor movement, the motor was used to move the screen imperceptibly back and forth, while still being audible. Panoramic renderings of all four scenes are shown in Fig. S9.

After conducting a Friedman test on the rankings, pair-wise follow up tests were conducted using the Wilcoxon signed rank test and corrected for multiple comparisons.

Monovision Presbyopes use monovision (and other treatments like bifocal or multifocal lenses) to extend the range of distances over which they experience clear vision. In monovision, a different power lens is placed in front of each eye, affording one eye clear vision for near distances and the other

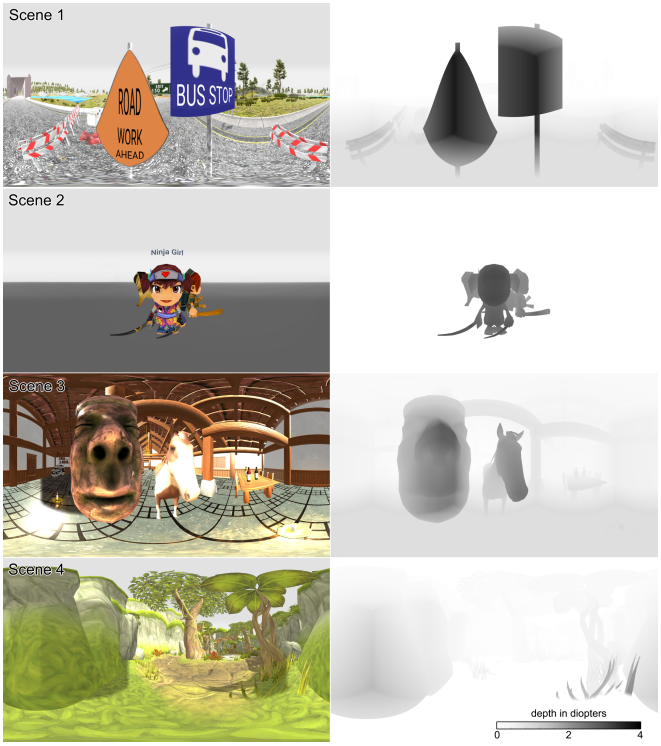

Fig. S9. Images of all four scenes used for the gaze-contingent preference test. These views show 360° panoramas using an equirectangular projection for both the color images (left column) and the depth maps (right column).

clear vision for far distances. Typically, the dominant eye is focused at 0 D (infinity), and the other eye at a nearer distance, usually at 1–2 D (0.5–1 m) (27, 28). The visual system then fuses the images together, with the acuity of the fused image being nearly equivalent to the better monocular acuity of each eye (29).

Monovision displays apply a similar principle to VR systems, but in this case with the goal of minimizing the vergence–accommodation conflict for non-presbyopes (4, 23, 24). A monovision-based VR system uses lenses of different focal power for each eye to place the virtual image of the display for each eye at a different distance (Fig. S6). Due to the fact that the accommodation for both eyes is linked and driven to the same distance (30), the hypothesis is that both eyes will accommodate to the virtual image plane closer to the distance that the user verges to, effectively mitigating the vergence–accommodation conflict one eye at a time. The downside is that one eye would always see a blurred image, but it is expected that the brain retains high frequency information from the eye with clearer vision for any condition (29). Such a scheme is attractive because it theoretically allows the user to accommodate to two planes, without requiring dynamic focal adjustments in the near-eye display.

As part of the study using the benchtop system, 16 users identified as not needing corrective eyewear (or wearing contacts) were tested in the monovision display mode (12 males, 4 females, age range 22–42 y old). When monovision correction is used for presybopia, the dominant eye is usually assigned to far focus. In our study, we always assigned the farther accommodative plane to the right eye, which is more commonly dominant (31, 32). For examining accommodative gain, we

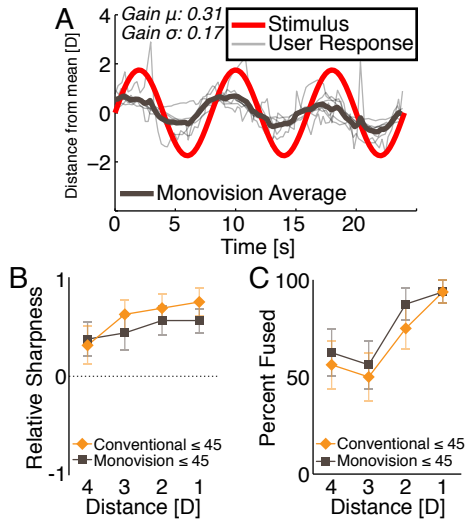

Fig. S10. Using the benchtop system to display a sinusoidal stimulus from 0.5–4 D (mean 2.25 D, amplitude 1.75 D), each user's accommodative response was recorded with monovision (*A*). The stimulus (red), each individual response (gray), and the average response (dark gray), are shown with the mean subtracted. These users also reported sharpness (*B*) and fusibility (*C*) with a stimulus displayed at four fixed depths from 1–4 D. The x-axis is reversed to show nearer distances to the left. Error bars indicate standard error.

only kept 7 measurements from the monovision condition due to dropped data points (Fig. S10*A*). The average accommodative gain of users in monovision was similar to the users in conventional mode aged 45 y old and younger (0.31 and 0.32, respectively). This difference was not statistically significant.

We see in Fig. S10 *B* and *C* that perceived sharpness and binocular fusion in monovision are comparable to the conventional condition. The user responses for monovision suggest that the target was slightly less sharp, and slightly more fusable. However, these differences were not statistically significant. Logistic regressions on the sharpness and fusibility reveal only a significant main effect of distance. The odds ratios for distance were 0.51 ($ci = 0.32$–0.79) and 0.17 ($ci = 0.067$–0.42), respectively ($ps < 0.01$).

The preference results also did no indicate substantial differences between monovision and conventional conditions. In this task, 6 of the 16 users chose monovision (not significantly different from chance). The results therefore do not suggest any improvement in perception or preference with monovision. Furthermore, the accommodative gains for monovision and conventional do not support the idea that users switch between accommodating to the two virtual image planes in monovision; the planes were 1.5 D apart, and that would correspond to a gain of 0.43. On the other hand, users were only given a few minutes to experience and evaluate monovision. It is possible that perceptual performance may be improved with extended experience of monovision in VR. For example, presbyopes who are prescribed monovision correction report a period of adjustment over days or even weeks (33). Interestingly, this subjective improvement over time does not seem to correspond to changes in objective visual measures (33, 34). Further studies will be necessary to investigate long-term effects of monovision-based VR/AR systems, both in terms of whether experience might improve perceptual performance, and whether there are negative effects of viewing monovision displays for non-presbyopes.

1. Banks MS, Hoffman DM, Kim J, Wetzstein G (2016) 3d displays. *Annual Review of Vision Science* 2(1):397–435.
2. Mauderer M, Conte S, Nacenta MA, Vishwanath D (2014) Depth perception with gaze-contingent depth of field. *ACM SIGCHI* pp. 217–226.
3. Maiello G, Chessa M, Solari F, Bex PJ (2015) The (in)effectiveness of simulated blur for depth perception in naturalistic images. *PLoS ONE* 10(10):1–15.
4. Konrad R, Cooper EA, Wetzstein G (2016) Novel Optical Configurations for Virtual Reality: Evaluating User Preference and Performance with Focus-tunable and Monovision Near-eye Displays. *Proc. SIGCHI* pp. 1211–1220.
5. Peli E, Hedges TR, Tang J, Landmann D (2012) 53.2: A binocular stereoscopic display system with coupled convergence and accommodation demands. *SID Digest* 32(1):1296–1299.
6. Schowengerdt B, Seibel E (2006) True 3-d scanned voxel displays using single or multiple light sources. *J. SID* 14(2):135–143.
7. Dolgoff E (1997) Real-depth imaging: a new 3d imaging technology with inexpensive direct-view (no glasses) video and other applications. *Proc. SPIE 3012* pp. 282–288.
8. Akeley K, Watt S, Girshick A, Banks M (2004) A stereo display prototype with multiple focal distances. *ACM Trans. Graph. (SIGGRAPH)* 23(3):804–813.
9. Rolland J, Krueger M, Goon A (2000) Multifocal planes head-mounted displays. *Applied Optics* 39(19):3209–3215.
10. Waldkirch M, Lukowicz P, Tröster G (2004) Multiple imaging technique for extending depth of focus in retinal displays. *Optics Express* 12(25).
11. Liu S, Cheng D, Hua H (2008) An optical see-through head mounted display with addressable focal planes in *Proc. ISMAR*. pp. 33–42.
12. Love GD et al. (2009) High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express* 17(18):15716–25.
13. Hu X, Hua H (2014) Design and assessment of a depth-fused multi-focal-plane display prototype. *Journal of Display Technology* 10(4):308–316.
14. Narain R et al. (2015) Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Trans. Graph. (SIGGRAPH)* 34(4).
15. Llull P et al. (2015) Design and optimization of a near-eye multifocal display system for augmented reality in *OSA Imaging and Applied Optics*. p. JTH3A.5.
16. Campbell FW (1957) The Depth of Field of the Human Eye. *Optica Acta: International Journal of Optics* 4:157–164.
17. MacKenzie KJ, Hoffman D, Watt SJ (2010) Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision* 10(8).
18. MacKenzie KJ, Dickson RA, Watt SJ (2012) Vergence and accommodation to multiple-image-plane stereoscopic displays: "real world" responses with practical image-plane separations? *Journal of Electronic Imaging* 21(1):011002–1–011002–8.
19. Lanman D, Luebke D (2013) Near-eye light field displays. *ACM Trans. Graph. (SIGGRAPH Asia)* 32(6):220:1–10.
20. Hua H, Javidi B (2014) A 3d integral imaging optical see-through head-mounted display. *Optics Express* 22(11):13484–13491.
21. Huang FC, Chen K, Wetzstein G (2015) The light field stereoscope: Immersive computer graphics via factored near-eye light field display with focus cues. *ACM Trans. Graph. (SIGGRAPH)* 34(4):60:1–12.
22. Campbell FW, Westheimer G (1960) Dynamics of accommodation responses of the human eye. *Journal of physiology* 151:285–295.
23. Johnson P et al. (2016) Dynamic lens and monovision 3D displays to improve viewer comfort. *Optics Express* 24:11808–11827.
24. Johnson PV et al. (2016) Assessing visual discomfort using dynamic lens and monovision displays in *Imaging and Applied Optics 2016*. (Optical Society of America), p. TT4A.1.
25. Pamplona VF, Mohan A, Oliveira MM, Raskar R (2010) Netra: Interactive display for estimating refractive errors and focal range. *ACM Trans. Graph.* 29(4):77:1–77:8.
26. Kuan PF, Huang B (2013) A simple and robust method for partially matched samples using the p-values pooling approach. *Statistics in Medicine* 32(19):3247–3259.
27. Legras R, Hornain V, Monot A, Chateau N (2001) Effect of induced anisometropia on binocular through-focus contrast sensitivity. *Optometry & Vision Science* 78(7):503–509.
28. Hayashi K, Yoshida M, Manabe S, Hayashi H (2011) Optimal amount of anisometropia for pseudophakic monovision. *Journal of Refractive Surgery* 27(5):332–338.
29. Zheleznyak L, Sabesan R, Oh JS, MacRae S, Yoon G (2013) Modified monovision with spherical aberration to improve presbyopic through-focus visual performance. *Investigative Ophthalmology & Visual Science* 54(5):3157.
30. Ball EA (1952) A study in consensual accommodation. *American journal of optometry and archives of American Academy of Optometry* 29(11):561–574.
31. Chaurasia BD, Mathur BBL (1976) Eyedness. *Cells Tissues Organs* 96(2):301–305.
32. Ehrenstein WH, Arnold-Schulz-Gahmen BE, Jaschinski W (2005) Eye preference within the context of binocular functions. *Graefe's Archive for Clinical and Experimental Ophthalmology* 243(9):926–932.
33. Collins M, Bruce A, Thompson B (1994) Adaptation to monovision. *International Contact Lens Clinic* 21(11):218 – 224.
34. Schor C, Carson M, Peterson G, Suzuki J, Erickson P (2016) Effects of interocular blur supression ability on monovision task performance. *Journal of the American Optometric Association* 60:188–192.