Novel Optical Configurations for Virtual Reality: Evaluating User Preference and Performance with Focus-tunable and Monovision Near-eye Displays

Robert Konrad
Stanford University

Emily A. Cooper
Dartmouth College

Gordon Wetzstein
Stanford University

ABSTRACT

Emerging virtual reality (VR) displays must overcome the prevalent issue of visual discomfort to provide high-quality and immersive user experiences. In particular, the mismatch between vergence and accommodation cues inherent to most stereoscopic displays has been a long standing challenge. In this paper, we evaluate several adaptive display modes afforded by focus-tunable optics or actuated displays that have the promise to mitigate visual discomfort caused by the vergence-accommodation conflict, and improve performance in VR environments. We also explore monovision as an unconventional mode that allows each eye of an observer to accommodate to a different distance. While this technique is common practice in ophthalmology, we are the first to report its effectiveness for VR applications with a custom built set up. We demonstrate that monovision and other focus-tunable display modes can provide better user experiences and improve user performance in terms of reaction times and accuracy, particularly for nearby simulated distances in VR.

Author Keywords

virtual reality; focus cues; user comfort; user performance

ACM Classification Keywords

H.1.2. User/Machine Systems: Human Factors; I.3.7 Three-Dimensional Graphics and Realism

INTRODUCTION AND MOTIVATION

Virtual and augmented reality (VR/AR) technologies have the potential to profoundly impact our society in the near future. With a wide range of applications in education, simulation and training, collaborative work, teleconferencing, scientific visualization, remote-controlled vehicles, telesurgery, basic vision research, entertainment, and phobia treatment, VR and AR provide a platform for immersive user experiences and unprecedented interaction techniques unmatched by other consumer electronics devices.

Recent advances in immersive near-eye display technologies promise consumer products to be within reach. Along with further reducing latency, increasing display resolution,

Figure 1. Focus-tunable near-eye display with different rendering modes. In the top panels, we illustrate via photographs how the images seen by each eye in a typical near-eye display are only focused properly when the eyes are focused at a relatively far distance (lower row). When the eyes focus near, to match objects simulated to be near, both eyes' images are blurry (upper row). Monovision is an alternative display mode where the lenses of the two eyes have different focal lengths, allowing for each eye to accommodate at a different distance. The bottom panels illustrate this mode. When the camera is focused relatively far, the left eye's image is out of focus and the right eye's image is sharp (lower row). When the camera is focused near, the relationship reverses. We asked whether this display mode could improve visual comfort and performance.

and providing compelling and interactive multi-modal experiences, the support for visually comfortable experiences may be one of the most important characteristics to determine the success of VR/AR technology. Current-generation
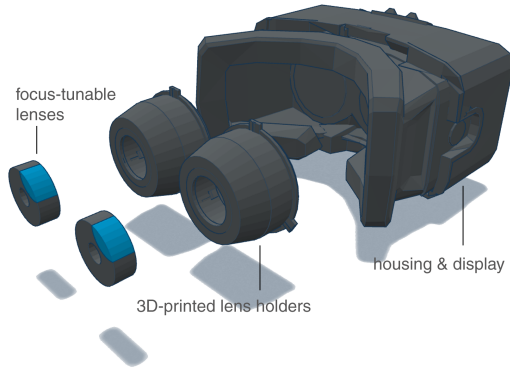
Figure 2. Focus-tunable near-eye display prototype configuration.

near-eye displays can simulate many of the depth cues typically available to the visual system, including both monocular and binocular cues (disparity and vergence). Unfortunately, focus cues (accommodation and retinal blur) are usually not supported, resulting in conflicting visual stimulation known as the vergence-accommodation conflict (Fig. 1, top panels) (see [21] for review). This conflict can cause visual discomfort and fatigue, and compromised visual clarity [37]. Using benchtop *multi-focal-plane displays*, it has been shown that correct or nearly correct focus cues that reduce this conflict can also improve stereoscopic correspondence matching [16], 3D shape perception [42], and discrimination of larger depth intervals [14].

Inspired by recent work on multi-focal-plane displays, we built a wearable VR display prototype with focus-tunable optics (programmable liquid lenses) and perceptually-driven rendering algorithms that allows the accommodative state of the observer's eyes to be adjusted adaptively and independently (see Figs. 1, 2). This system can be used to generate correct or nearly-correct focus cues via time-multiplexed or adaptive focus image presentation, or via *monovision*. In monovision, an optical augmentation that is currently a treatment option for patients who lose the ability to focus, the two eyes observe the world through two lenses with different focal lengths. For some wearable display applications, monovision may be an effective tool to counteract the vergence-accommodation conflict.

Contributions to Human-Computer Interaction
While the ideas of gaze-contingent retinal focus and monovision are not new, our work makes the following contributions that advance the field towards integrating these improvements into standard VR applications:

- We quantify the perceptual benefits of multiple different focus-tunable display modes.

- We demonstrate and evaluate monovision for mitigating the vergence-accommodation conflict in VR displays.

- We directly compare monovision and focus-tunable modes w.r.t. user preference and performance.

- We examine factors that are relevant to typical HCI applications, such as reaction time and perceptual accuracy.

Our user studies advance our understanding of human factors in VR and suggest that monovision is a viable and low-cost path to mitigate the vergence-accommodation conflict. A system capable of producing both monovision and adaptive focus can be implemented with focus-tunable optics, as done in our studies, or with actuated displays. Such flexible systems can provide a valuable tool for HCI and graphics research. However, low-cost monovision setups can also be implemented using static lenses with different focal lenths or using two staggered screens.

However, it is also important to note that the current prototype has several limitations, most importantly a limited FOV and a lack of gaze tracking. In the current study, we employed a somewhat constrained setting with the field of view afforded by our prototype, and we encouraged users to fixate on one target at a time. Furthermore, the potential long-term effects of the monovision configuration on comfort in individuals with normal vision are unknown. Nevertheless, the insights we gained will directly benefit future HMDs as technology advances to overcome these limitations.

RELATED WORK

Focus-supporting Near-eye Displays
Sutherland's vision of electronic head mounted displays [40] has become practical and affordable with recent developments in the consumer electronics industry. However, providing accurate focus cues remains a major challenge for these types of displays. Possible options include *volumetric displays* that mechanically [10, 20] scan a projection surface to create a 3D display volume within which an observer can accommodate. Unfortunately, the form factor of these devices is not suitable for wearable displays. *Multi-focal-plane displays* approximate such a volume using a few virtual display planes that are generated by beam splitters [7, 1] or focus-tunable optics [34, 41, 17, 24, 27, 36, 25, 32]. Two possible ways to create focus cues with focus-tunable optics have been described: temporal multiplexing and gaze-contingent display. For temporal multiplexing, the focus is periodically swept through the volume and synchronized with the display update. However, display refresh rates beyond those offered by current-generation microdisplays are usually required and flicker may be perceived. Gaze-contingency can be achieved using eye tracking (e.g. [35]) by either estimating the vergence angle of the eyes or by analyzing the depth buffer of the displayed images at the gaze-tracked location. An example *gaze-contingent focus display* was implemented by Sugihara and Miyasato [39]: two monitors are physically moved towards or away from the observer based on their gaze direction. Although Peli [33] outlines prior work that proposes the idea of gaze-contingent focus, we are not aware of anyone having reported a practical gaze-contingent, focus-tunable display prototype yet, although this would be an obvious next step.

Monovision
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. Currently, there is no treatment available for presbyopia that restores the

eyes' ability to adaptively focus. As such, the goal of treatments such as surgery, bifocals, and monovision is to augment the presbyopic patient's optics in the way that restores the range of distances over which they can experience clear vision [13, 11]. To accomplish this goal, the monovision treatment method involves focusing one eye at a fixed near distance, and the other eye at a fixed far distance. This is accomplished via contact lenses, spectacles, or surgery. Ideally, this correction allows the patient to experience clear vision over a range of distances via whichever eye is closest to the correct focus (see [9] for a review). However, the differing focal distances of the two eyes means that one eye is necessarily out of focus at any given distance, so patients with monovision can experience overall decreases in acuity and stereoscopic depth perception (likely because the two eyes' images are dissimilar) [2, 13]. Nonetheless, patient satisfaction after monovision correction tends to be quite high [9]. Thus, it seems possible that a technique similar to monovision might be able to ameliorate some of the perceptual and visual-comfort consequences of the vergence-accommodation conflict in near-eye displays, by essentially minimizing the conflict in one eye at a time. Indeed, the idea of exotic optical configurations for near-eye displays, including monovision and bifocal lenses, was first described by Marran and Schor [30]. The idea for using monovision for this application has continued to be of interest [33] and benchtop implementations were proposed [26]. An evaluation of these ideas had not been reported when we conducted our work; however, we recently because aware of a concurrent study [19]. Our display prototype enabled us to assess the effectiveness of a monovision optics configuration, along with a variety of other focus-tunable display modes.

Retinal Blur Rendering

In addition to the prior work investigating the perceptual improvements associated with benchtop multi-focal-plane displays (which include both accommodative and retinal blur cues) summarized in the introduction, other researchers have investigated the perceptual effects of *gaze-contingent retinal blur* rendering on its own. Because gaze-contingent retinal blur rendering only requires a gaze-tracker and fast/realistic blur rendering techniques—no specialized optics are needed—it is useful to know if this type of display mode on its own offers improvements over standard displays. Several previous studies have examined the effect of this rendering technique on visual experience and performance with benchtop displays [5, 15, 28, 31, 8]. In these studies, gaze-tracking and estimated or ground-truth depth maps were used to adaptively update the depth of field of an image depending on the distance of the object that the participants were fixating. Several studies reported improvements in subjective viewing experience [15, 31], however, the results for performance improvements on a variety of visual tasks were more mixed [5, 31]. One study showed that combining this technique with stereo display significantly decreased the time needed for participants to achieve binocular fusion under some conditions [28]. Given the mixture of results, it seemed worthwhile to assess an implementation of retinal blur rendering in the current near-eye display system.

Field of view	$26°(H) \times 26°(V)$
Resolution	488×488 px
Accommodation range	10.5–∞ cm
Latency	11.5 ms rendering
	15 ms focus adjustment

Table 1. Overview of focus-tunable HMD characteristics.

Figure 3. Photograph of HMD prototype and focus-tunable lens.

Light Field Displays

Recently, microlens-based near-eye light field displays have been shown to support focus cues [23], albeit at low spatial resolution. Building on emerging compressive light field display technology [22, 43, 29], Huang et al. [18] recently demonstrated the first high-resolution near-eye light field display for VR applications. Whereas Huang's work achieves focus cues at moderate spatial resolution without temporal multiplexing, eye tracking, or focus-tunable optics, the required light field factorizations place a high demand on computation. Furthermore, reduced light throughput and diffraction blur of stacked liquid crystal displays are problematic for see-through AR display modes. The focus-tunable display modes evaluated in this paper are complimentary to recent developments in near-eye light field displays. By exploring focus-tunable stereoscopic displays, we hope to provide new insights and quantify the advantages of focus-tunable displays for both AR and VR applications. Finally, we anticipate the proposed techniques to enhance visual comfort also in light field displays, for example by optimizing the perceived depth of field in possible future implementations of focus-tunable near-eye light field displays via monovision.

FOCUS-TUNABLE NEAR-EYE DISPLAY SYSTEM

Hardware

Building on commodity electronics and optics components, we built a focus-tunable display based on the Oculus Rift DK2. The original lenses are replaced by two independently controllable Optotune EL-10-30-VIS-LD focus-tunable liquid lenses, driven by an Optotune Lens Driver 4 (see Fig. 3). Each of these plano-convex lenses has a diameter of 10 mm and they are tunable to a focal length range of 52–140 mm (19.23–7.14 D) throughout the visible spectrum. The lenses are mounted at a distance of 6 cm from the display panel. In our system this corresponds to an accommodation range of 10.5–∞ cm. The response time is approximately 15 ms at room temperature. The update of the lens focal state and display were synchronized. The focus tunable lenses are connected to a host computer (Dell XPS 8700 workstation, Intel

378 i7 CPU, 8 GB RAM, NVIDIA GTX 690 GPU) via USB and all rendering software is written in C++ and OpenGL.

The main limitation of the prototype are the small apertures of the focus-tunable lenses, which severely restrict the observed field of view and also crop most of the screen. When the lenses are set to a focal length of 8.5 cm the field of view is approximately 26°, which corresponds to a 488 × 488 px viewing area. The field of view changes by 1° within the focal length range. Note that the field of view limitation is specific to of our dynamically-switchable prototype; large fields of view with any of the discussed display modes on their own can be easily achieved with fixed-focus lenses.

Software
All hardware, including the display, the internal measurement unit (IMU), and the focus-tunable lenses are controlled from a C++ program. The render engine is embedded in that code; we use OpenGL with a physically plausible approximation of retinal blur / depth of field effect implemented in a fragment shader. We implemented an approximation of retinal blur via adaptive depth of field (DOF) rendering.

To perform the rendering, we used image space with the aid of a depth map. As explained in [4], there are several criteria that image space methods would ideally satisfy to simulate DOF blur: choice of point spread function (PSF), lack of intensity leakage, lack of depth discontinuity artifacts, proper simulation of partial occlusions, and high performance. However, no post-processing is able to achieve all of these simultaneously. In the experiments run, only a portion of the above were necessary. Depth discontinuity artifacts never occurred in the foveal region because the in-focus target was never occluded, and although our algorithm could handle such artifacts, they were not prevalent. Therefore, we focused on implementing two criteria well: lack of intensity leakage (a blurred background never blurs on top of an in-focus foreground) and high performance (critical for VR applications).

We followed the method in [44] which uses accurate circles of confusion to simulate DOF effects by adjusting weights of a non-linear filter. In our prototype system, we render two 960 × 1080 px scenes to one framebuffer for each eye and then apply the retinal blur to the whole framebuffer. The process of rendering the scenes and applying the blur takes 11.5 ms (86.9 FPS) on average across the three scenes that we tested. On average the retinal blur takes 7.67 ms to render. Note that focus adjustment in the lens can be triggered prior to applying the depth of field shader, which minimizes overall latency.

EVALUATION
We conducted three user studies to assess the effects of the focus-tunable display system and adaptive DOF rendering on VR applications. The goal of these studies was (1) to determine if users have a *preference* for engaging with VR environments using different display modes, and (2) to determine whether user *performance* on tasks within VR environments is affected by different display modes. The research protocol was approved by the Institutional Review Board at Stanford University and all participants gave informed consent.

Participants: Twelve users participated in all three studies. Participants were recruited from the university population (age range 18-28, 3 females). All participants reported having normal or corrected-to-normal vision, and passed a standard Randot stereo vision test. Two volunteers were excluded from participating because they did not pass the stereo vision test.

Display modes: For each study, we compared 5 modes:

- *Normal* mode: equivalent to standard usage of the Oculus, the focus-tunable lenses were disabled (but present) and VR environments were rendered with infinite DOF

- *Adaptive DOF* mode: the focus-tunable lenses were disabled, but VR environments were rendered with adaptive DOF that placed the main point-of-interest in the scene at the center of the DOF

- *Adaptive Focus* mode: the focus-tunable lenses were engaged and focused adaptively to the distance of the main point-of-interest, and VR environments were rendered with infinite DOF

- *Adaptive Focus + DOF* mode: the focus-tunable lenses were engaged and focused adaptively to the distance of the main point-of-interest, and adaptive DOF was used to also put this point at the center of the DOF

- *Monovision* mode: the focus-tunable lenses were fixed at two different focal distances for the two eyes, 2 diopters for the left eye and infinity for the right eye, and adaptive DOF rendering was enabled

Throughout the descriptions of these user studies, we will refer to the simulated distances of objects in the VR environment in terms of diopters (D), which are measured in inverse meters, as well as in meters (m). Also, note that our prototype does not include eye-tracking. Thus, in each user study, the participants were instructed as to where to look in the scene via the presentation of specific targets.

User Preferences Study
Methods: In this study, we asked users to visually explore a VR environment and rank the overall quality of their experience for different display modes. Users were fitted with the display system, and then allowed to switch freely through the five display modes for a single VR environment; the screen was not blanked when switching between display modes. The environment consisted of a static scene with a single target floating dynamically within the scene (Figure 4, left panel). For the four adaptive conditions (*Adaptive DOF*, *Adaptive Focus*, *Adaptive Focus + DOF*, and *Monovision*), this target was treated as the main point-of-interest for which the DOF and/or focus were set. The simulated distance of the target ranged from 3.33 to 0.25 D (0.3 to 4 m). For the non-adaptive, fixed focus reference mode, the in-focus accommodative distance was 1.3 m (0.77 D)[1]. For all display modes using a rendered depth of field, we assume a pupil diameter of 4 mm. The users were asked to rank the modes from 1-5 in terms of their general viewing experience, with 1 being the best score. As the users set their rankings, a number on the floating target was updated to indicate the ranking of the current mode. The

---

[1]Similar to DK2; see Oculus Best Practices Guide

Figure 4. Example screenshots from the user studies. For the user preference study (left), we ask users to fixate on a target moving periodically in depth within a static scene. The target shows a number with the user's ranking of the current display mode, while the character is randomly associated with the display mode. For the visual clarity (center) and depth judgment (right) tasks, we display small targets that are magnified in the insets.

initial display mode and the order through which they were switched was randomized for each session, and users were encouraged to view each mode at least once before starting to rank them. They were also encouraged to focus on the floating rectangle during the task, although this instruction was not strictly enforced. They repeated their ranking three times for three different VR environments.

Results: On average, the users ranked the normal mode the lowest, and they preferred the display modes with adaptive focus. Figure 5 shows the results for each condition. The bar heights indicate the mean ranking (and standard error) across subjects. Individual circles show the mean rankings for each individual user. The rankings for the Adaptive Focus and Adaptive Focus + DOF conditions were highest (2.2 and 2.4, respectively). Individual user data reveal that these two modes were often selected as either the first or second choice. Normal mode and Adaptive DOF alone had the lowest rankings (3.9 and 3.5, respectively). The Monovision mode fell in the middle (3.2).

To test for statistical significance, we performed a Friedman test and follow up multiple comparison tests (using the Tukey-Kramer method) on the mean ranking, after averaging the repeated trials for each subject. These tests determined if display mode had a significant effect on user mean rankings, and which specific display modes resulted in significant improvements above Normal mode, respectively [38]. The Friedman test indicated a significant effect of display mode ($p < 0.01$). A followup multiple comparison test with a significance threshold of $p < 0.05$ showed that the rankings for Adaptive Focus and Adaptive Focus + DOF were significantly better than Normal mode, but the other display modes were not.

These results suggest that manipulating focal distances adaptively can improve subjective viewing experience. While the two other display modes (Adaptive DOF and Monovision) did not significantly improve user rankings, the results indicate that they also did not degrade viewing experience. Thus, if either of these display modes confers a performance advantage, they might be able to do so without any cost to subjective experience.

After each user completed this experiment, we also conducted an informal debriefing interview. Users consistently commented that their rankings were affected by the "blurring of
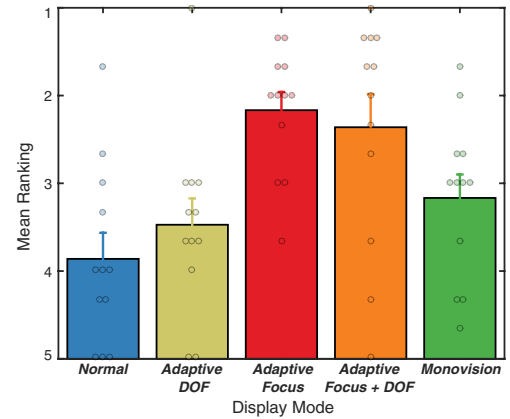

Figure 5. User preference study results. Bar heights indicate the mean ranking for each display mode across all users. A ranking of one is for the best experience, and five the worst. Note that one is at the top of the ordinate axis and five is at the bottom. Error bars are the between-subjects standard error of the mean. Circles indicate the mean ranking for each individual user across three repetitions.

the background" and not being able to fuse the floating square at close distances. Also, many commented that the lenses tended to get warm.

User Performance Studies

Methods: In this study, we asked if user performance on two different tasks in a VR environment was affected by the different display modes. In the *Visual Clarity* task, users identified the orientation of a small letter "E" placed on a target within a VR scene. In the *Depth Judgment* task, users made a relative depth judgment between two targets within the scene. The procedure and stimulus for each task were almost identical. At the start of each trial, the user was instructed to fixate an initial target, a textured gray square, presented in the center of the visual field. This target could appear at one of three simulated viewing distances: 0.4, 1, and 2 D (2.5, 1, and 0.5 m). Once they were fixating, the user initiated the trial with a key press. After a random delay (uniformly distributed between 0 and 1 seconds), the first target disappeared and a second target appeared at one of five simulated depths relative to the viewing distance. In diopters, these relative distances were the same for each viewing distance: -0.5, -0.2, 0, 0.2, and 0.5 D. That is, sometimes the second target appeared in the same distance plane as the first, sometimes it was closer

Visual Clarity Task

(a)
— *Reaction Times* —

(c)
— *Accuracy* —

Depth Judgment Task

(b)
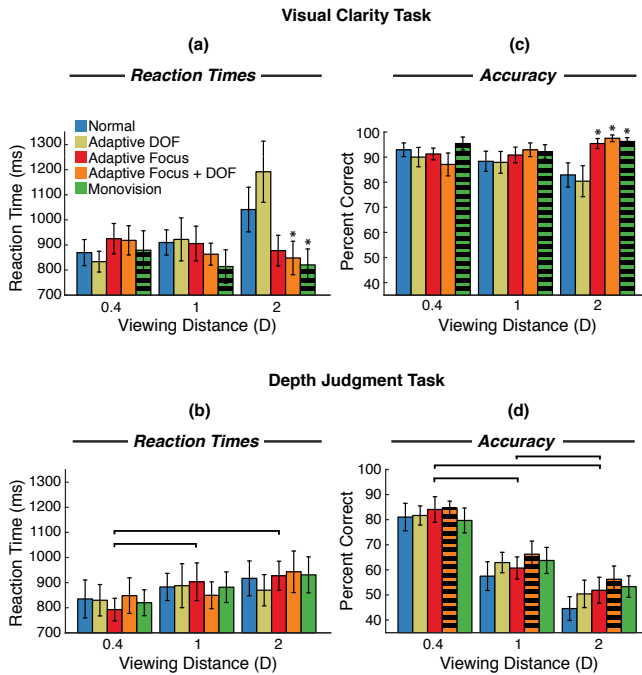— *Reaction Times* —

(d)
— *Accuracy* —

Figure 6. User performance study results. For each performance task, panels (a) and (b) show the results for reaction times, and (c) and (d) show results for accuracy. Each bar indicates the results for one of the display modes, averaged across users (see legend). Bars are grouped and plotted separately by viewing distance. Note that the data for each user from each viewing distance were initially averaged over the different target depths. To reflect the ANOVA, the farthest target depth for viewing distances of 1 and 2 D are excluded, so that the same set of four target depths are included in each group. For data from all individual targets (see Fig. 7). Error bars indicate standard errors across subjects.

to the user, and sometimes it was farther away. Obviously, the relative depth in meters depended on the viewing distance of the first target. For the farthest viewing distance (0.4 D), the relative distance of -0.5 D was excluded because this was an impossible object distance. The second target appeared randomly within a circle indicated by the intersection of a 2° cone and that distance plane. In both experiments, all of the targets subtended the same visual angle regardless of the simulated distance.

In the *Visual Clarity* task, the second target had a small letter "E" (6 pixels in height, or 22.59 arcminutes), which was facing upwards, downwards, leftwards, or rightwards (Figure 4, middle panel). Users responded with a key press to indicate the letter's orientation. In the *Depth Judgment* task, the second target instead simply had a number "2" on it, and the users responded whether this target was closer, farther, or at the same distance as the first target (Figure 4, right panel). In both tasks, the users were instructed to make their response as quickly as possible. The two tasks were run in two separate sessions. Within each task, each combination of viewing distance and relative distance was repeated five times for each user, and the order of presentation was randomized. We recorded the reaction times (time between when the second target appeared and when the user logged their response) and accuracy (whether or not they responded correctly).

Analysis: To examine differences in reaction time, for each user we calculated the median reaction time over all repetitions of a given trial type (display mode × viewing distance × target relative depth). We used the median in order to reduce the impact of outlier trials with very long reaction times. To examine differences in accuracy, we calculated the percentage of trials in which each subject responded correctly for each trial type. For each of the user performance studies, a 5 × 3 repeated measures analysis of variance (ANOVA) was performed to test for statistically significant main effects and interactions of two fixed factors (Display Mode and Viewing Distance). Display Mode had five levels (Normal, Adaptive DOF, Adaptive Focus, Adaptive Focus + DOF, and Monovision) and Viewing Distance had three levels (0.4D, 1D, and 2D). To avoid having a large number of factors, we did not examine the effect of each individual target depth with the ANOVA. Thus, this analysis was performed on data collapsed across target depths for each viewing distance, with the farthest target depth (-0.5D) excluded from analysis because it was not present at all viewing distances. We also calculated generalized eta squared (GES) measures of effect size [3]. Follow-up t-tests were performed using all data points to examine the statistically significant effects from the ANOVA, and p-values were corrected for multiple comparisons using the Holm method. These t-tests were further limited to three different planned comparisons which were focused on examining effects of viewing distance, and differences between Normal mode and the other display modes: (1) comparing reaction time/accuracy for each viewing distance, (2) comparing each augmented display mode to Normal mode, (3) comparing each augmented display mode to Normal mode for each viewing distance separately.

Results

The analyses suggest that adaptive focus and monovision display modes can result in better user performance under a number of conditions. The results for both performance studies are summarized in Figure 6 and Figure 7. Figure 6 shows the average reaction time and accuracy across users for each combination of display mode and viewing distance. The results for the Normal display mode are shown in blue. Display modes that were significantly different from normal (main effects), are indicated by striped bars, interactions between display mode and viewing distance are indicated by asterisks above the bars that were significantly different from the normal mode, and main effects of viewing distance are indicated by horizontal lines. Figure 7 shows the same data, separated out for each individual target location, and thus reveals under which conditions some of the largest differences between display modes appear. For simplicity, in Figure 7 we show the mean reaction times and accuracy data for each mode relative to the Normal mode. This way of plotting makes it clear whether the added features (DOF, focus-tunable lenses, monovision) increased or decreased reaction times. Points that lie near zero indicate that reaction times and accuracy were similar to those in Normal mode. Points above zero (white portion) indicate faster reaction times and better accuracy, points below zero (gray portion) indicate slower reaction times and worse accuracy. For both reaction times
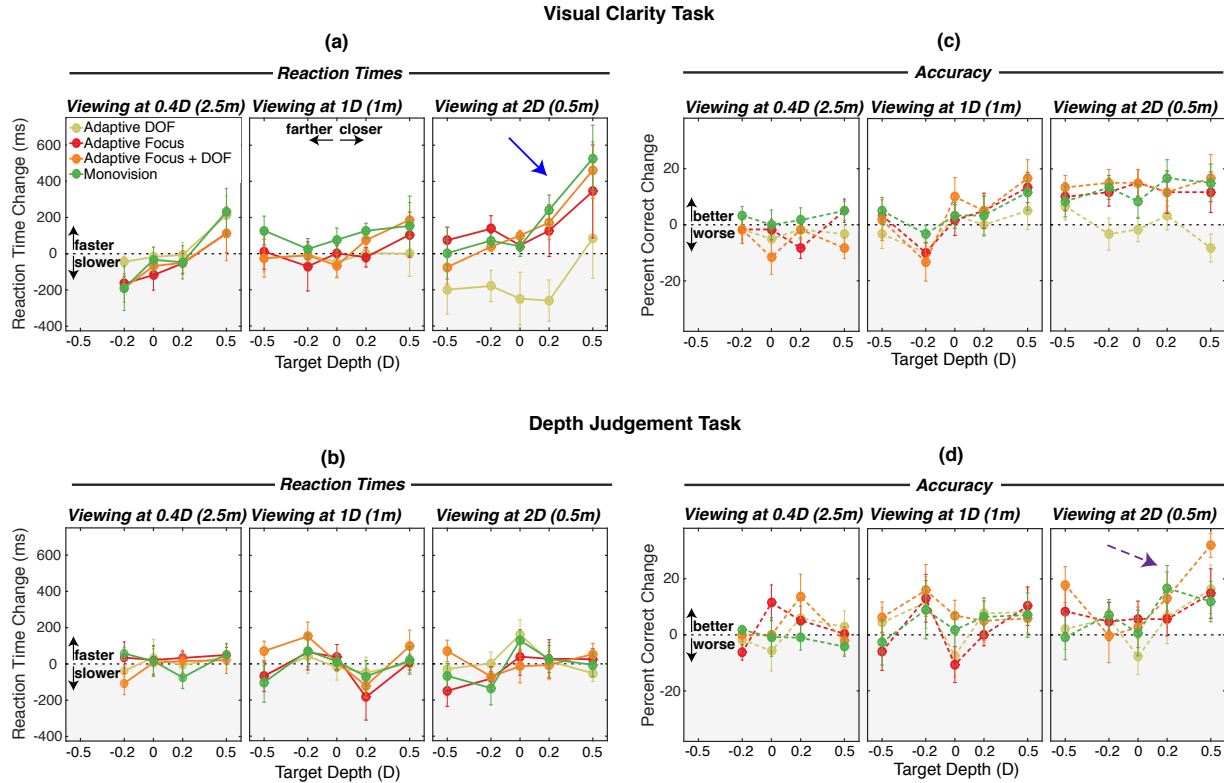
Visual Clarity Task

(a) Reaction Times

Viewing at 0.4D (2.5m) · Viewing at 1D (1m) · Viewing at 2D (0.5m)

Legend: Adaptive DOF · Adaptive Focus · Adaptive Focus + DOF · Monovision

(c) Accuracy

Viewing at 0.4D (2.5m) · Viewing at 1D (1m) · Viewing at 2D (0.5m)

Depth Judgement Task

(b) Reaction Times

Viewing at 0.4D (2.5m) · Viewing at 1D (1m) · Viewing at 2D (0.5m)

(d) Accuracy

Viewing at 0.4D (2.5m) · Viewing at 1D (1m) · Viewing at 2D (0.5m)

Figure 7. User performance study results for all targets. Panels (a) and (b) show the results for reaction times, and (c) and (d) show results for accuracy. Each colored line shows the results for a different display mode (see legend). Results for Normal mode are not shown, because the results for each of the other modes were normalized by taking the difference from the reaction time or percent correct for Normal mode. These differences were calculated for each user, and the mean and standard errors of these differences across users are shown as circles and vertical bars.

and accuracy, differences were calculated by taking the difference between the Normal mode data and the other display modes, such that differences are represented as changes in reaction time in ms, and changes in percent correct in percentage points, rather than ratios. Each column shows the results for all target depths for a single simulated viewing distance.

Visual Clarity/Reaction Times: Across all trial types, the average reaction time in the Visual Clarity task was 913 ms. Particularly for the nearest viewing distance (2 D, 0.5 m), several display modes resulted in faster reaction times than the Normal mode. The ANOVA showed a significant main effect of Display Mode, $F(4,44) = 3.8$, $p < 0.01$, and a significant interaction between Display Mode and Viewing Distance, $F(8,88) = 6.4$, $p < 0.001$. The sizes of these effects (GES) were 0.04 and 0.09, respectively. These and most other effects in these data were in the range of small to medium [3]. Follow up t-tests showed that the reaction times for the Monovision mode across all viewing distances were significantly faster than for Normal mode ($p < 0.05$), and that at the nearest viewing distance (2 D), the reaction times for Monovision and Adaptive Focus + DOF were also significantly faster ($ps < 0.05$) (See Figure 6(a)). Figure 7(a) shows that the display modes with Adaptive Focus and Monovision resulted in faster reaction times than the Normal mode particularly when the simulated distances were relatively close to the user (2 D/0.5 m or less; blue arrow in right panel). This advan-

tage was largely due to the fact that reaction times in the Normal mode slowed down at these distances, whereas reaction times in the adaptive focus conditions remained relatively stable. In Normal display mode, the virtual screen is optically placed at 0.77D, therefore simulating a viewing distance at 2 D results in the strongest vergence-accommodation conflict (targets appearing from 1.5 D to 2.5 D). The larger conflict may lead to longer image fusion/response times, hence the increased baseline reaction times. At the 2 D viewing distance, the Adaptive DOF mode actually tended to result in slower reaction times than Normal mode (yellow line), however, this slow down did not reach statistical significance.

Depth Judgment/Reaction Times: Unlike the Visual Clarity task, reaction times in the Depth Judgment task did not vary substantially as a function of display mode. Across all trial types, the average reaction time was 884 ms. The ANOVA showed a significant main effect of Viewing Distance, $F(2,22) = 7.3$, $p < 0.01$, $GES = 0.03$, but no effect of Display Mode and no interaction. Follow up t-tests showed that pair-wise comparisons between the farthest and two closer viewing distances were significant ($ps < 0.01$), indicating that reaction times tended to slow down for closer targets, but that there was no significant effect on reaction times of the different Display Modes (See Figure 6(b) and Figure 7(b)).

Visual Clarity/Accuracy: The reaction time data suggest that the adaptive focus display modes offer an advantage in terms

of being able to react quickly to nearby stimuli within a VR environment. However, if this increase in speed was also associated with a decrease in performance accuracy, it might be undesirable. Thus, we also asked whether users' response accuracy was affected by the different display modes. Across all trial types in the Visual Clarity task, average accuracy was 91% (chance performance would be 25%). This is not surprising, because all users reported having normal vision, and the "E" was large enough that it was clearly visible if the eyes were properly fixated and focused. Thus, incorrect trials likely result from users choosing to respond before the target was clearly focused, in order to comply with the instruction to respond as quickly as possible. Interestingly, user accuracy did not decrease for the display modes with faster reaction times, but actually increased slightly (See Figure 6(c) and Figure 7(c)). As with the reaction times for this task, the ANOVA showed a significant main effect of Display Mode, $F(4,44) = 3.6$, $p < 0.05$, $GES = 0.07$, and an interaction between Display Mode and Viewing Distance, $F(8,88) = 4.9$, $p < 0.001$, $GES = 0.08$. Follow up t-tests showed that the accuracy results followed a similar pattern to the reaction times results: Monovision had better accuracy across all distances ($p < 0.01$), and at the nearest viewing distance (2D), the accuracy for Monovision, Adaptive Focus + DOF, and (in addition) Adaptive Focus display modes were also significantly more accurate ($ps < 0.01$).

Depth Judgment/Accuracy: In the Depth Judgment task, the overall performance was worse than the Visual Clarity task (67% correct on average across all trial types, chance was 33%). Surprisingly, despite the fact that reaction times were similar across display modes in this task, we did observe a small but statistically significant improvement in accuracy with one adaptive focus display mode (See Figure 6(d)). The ANOVA showed a significant and large main effect of Viewing Distance, $F(2,22) = 44.7$, $p < 0.001$, $GES = 0.39$ and a main effect of Display Mode, $F(4,44) = 2.7$, $p < 0.05$, $GES = 0.02$. Follow up t-tests showed that all pair-wise comparisons between viewing distances were significant, indicating worse accuracy for closer targets ($ps < 0.001$). There was also a significant improvement in accuracy relative to Normal mode for the Adaptive Focus + DOF mode ($p < 0.01$). An examination of the individual target data (Figure 7(d)) reveals a similar pattern to that observed in the Visual Clarity Task reaction times: the most improvement in performance with the adaptive focus and monovision modes occurs when the simulated distances are relatively near to the observer (purple arrow).

DISCUSSION

In summary, we introduce and evaluate monovision for mitigating the vergence-accommodation conflict in VR displays, we quantify the perceptual benefits of newly available focus-tunable display modes, and we directly compare monovision and focus-tunable modes w.r.t. user preference and performance. Our user studies reveal that monovision and adaptive focus can offer higher-quality VR experiences and better user performance than conventional display modes, particularly for objects and tasks performed at nearby depths. Gaze-contingent depth of field rendering on its own, as proposed in previous work, may not offer the same benefits as monovision or gaze-contingent focus.

Interpretation of User Study Results
The results of our user studies support the idea that focus adaptability and monovision can improve user experience and performance with stereoscopic near-eye displays. We can conclude that particularly in cases where it is desirable to simulate object distances that are relatively close to the user (less than 0.5 m), focus adaptable technology provides desirable benefits. This makes sense, because without adaptable lenses, the fixed focal distance of the display we used for testing is 1.3 m. Thus, without the adaptable lenses, simulated near distances produce a large vergence-accommodation conflict.

Somewhat surprisingly, the improvements of the adaptive refocusing conditions (Adaptive Focus and Adaptive Focus + DOF) were nearly matched, and in some cases surpassed, by the Monovision display mode. This is an intriguing finding, because while implementing eye-tracking to guide adaptive focusing in near-eye displays with focus tunable optics poses substantial technical challenges, monovision can easily and inexpensively be implemented by simply placing two fixed lenses of different powers. What we do not know is whether any potential long-term discomfort associated with decoupling the accommodative distances of the two eyes in monovision would be any greater than the discomfort that is already associated with the vergence-accommodation conflict in near-eye displays. In the current work, the user studies were relatively short in duration, so this will be an interesting avenue for future work.

Lastly, it is unclear why our implementation of the Adaptive DOF display mode alone reduced reaction times, in the context of prior work showing that retinal blur cues can aid binocular fusion [28]. Although neither of our tasks explicitly tested binocular fusion, it would be reasonable to assume that if time to fusion was substantially faster, the users might have responded faster in the Visual Clarity and/or Depth Judgment task. One possibility is that we used computer generated scenes, which are likely less cluttered and detailed than the natural images used the prior study.

Alternative Hardware Implementations
The insights drawn from our experiments can serve as guidelines for the design of future near-eye displays. If the user's gaze direction can be either measured (i.e. using gaze tracking) or otherwise predicted, gaze-contingent focus is the most natural viewing condition and should be used if possible. The primary benefit of monovision, as an alternative to gaze-contingent focus, is its simplicity. It can be implemented using focus-tunable lenses, as demonstrated in our prototype, using actuated screens in the HMD, or using conventional, fixed-focal length lenses. The latter option can be readily implemented with any existing HMD - simply mount two different lenses in the HMD and adjust the stereo rendering. We chose focus-tunable lenses for our prototype, because it allows us to quickly switch between different rendering modes to quantify and compare them with user studies.

Additional Limitations

Focus-tunable lenses require a fair amount of power and, unfortunately, much of this power is dissipated as heat. The lenses tend to heat up particularly when pushed to low focal lengths. According to the manufacturer, when the lenses heat up the fluid inside of the lenses expands in volume. Therefore the focal length of the lenses decreases by as much as 0.67 D per 10°C temperature increase. We accounted for this by calibrating the interaction between current, focal length, and temperature of the lenses. On average, there was a 0.08 D error over the range of focal lengths used, with the highest error being 0.17 D. However, we know that the accommodation system of the human visual system is not perfect. People can naturally have fairly large accommodative lags during natural viewing situations (i.e. they do not accommodate enough to a near stimulus) [12] and the accommodative state of the eye can fluctuate by several 10ths of a diopter [6]. Thus, it is currently unclear how precisely the focus tunable lenses need to be calibrated in order to provide perceptual benefits.

Future Work
In the near future, we would like to run more user studies that evaluate these new display modes for VR applications. Our current study included a typical user sample size (see [28, 31, 15]), however, it would be advantageous in the future to increase the sample size, the variety of tasks, and the duration of time during which the prototype is worn. Furthermore, for the monovision display made, the choice of focal length could be varied to determine the best accommodative state for VR. With a setting of 2 D, we follow typical settings reported in the clinical literature [2, 9], but other settings are certainly possible. Testing monovision for see-through augmented reality applications would be interesting, because in that setting the real world provides natural focus cues that are optically overlaid with the virtual image and may make monovision display modes even more comfortable and interesting. In the long run, the effectiveness of monovision should be tested as a function that varies over the entire visual field, such that unconventional fixed optics (e.g., bi-focal and multi-focal lenses) can be optimized for spatially-varying, wide field of view monovision display modes in VR and AR.

CONCLUSION
Emerging near-eye displays have the potential to profoundly impact our society by providing a fundamentally new tool for education, collaborative work, scientific visualization, remote-controlled vehicles, training and simulation, and surgical training. Providing visually comfortable experiences with these displays, however, is a long standing challenge. With this work, we provide possible ways to decrease visual discomfort using focus tunable display modes and monovision. We have evaluated monovision and demonstrated that it can help users to detect targets more accurately and in less time than what is possible with conventional stereoscopic HMDs, without sacrificing comfortable viewing experiences. We believe our technology to be a crucial advance towards more immersive and comfortable VR experiences.

REFERENCES
1. K. Akeley, S.J Watt, A.R Girshick, and M.S Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Trans. Graph.* 23, 3 (2004), 804–813.

2. A. Back. 1995. Factors Influencing Success and Failure in Monovision. *Int. Contact Lens Clinic* 22, 7 (1995), 165–72.

3. R. Bakeman. 2005. Recommended effect size statistics for repeated measures designs. *Behavior research methods* 37, 3 (2005), 379–384.

4. B. A. Barsky and T. J. Kosloff. 2008. Algorithms for Rendering Depth of Field Effects in Computer Graphics. In *Proc. Int. Conf. on Computers*. 999–1010.

5. J. P. Brooker and P.M. Sharkey. 2001. Operator performance evaluation of controlled depth of field in a stereographically displayed virtual environment. *Proc. SPIE* 4297 (2001), 408–417.

6. WN Charman and G. Heron. 1988. Fluctuations in accommodation: a review. *Ophthalmic & physiological optics* 8, 2 (1988), 153–164.

7. E. Dolgoff. 1997. Real-Depth imaging: a new 3D imaging technology with inexpensive direct-view (no glasses) video and other applications. *Proc. SPIE 3012* (1997), 282–288.

8. A. T. Duchowski, D. H. House, J. Gestring, R. I. Wang, K. Krejtz, I. Krejtz, R. Mantiuk, and B. Bazyluk. 2014. Reducing Visual Discomfort of 3D Stereoscopic Displays with Gaze-contingent Depth-of-field. In *ACM Symposium on Applied Perception*. 39–46.

9. B. J. Evans. 2007. Monovision: a review. *Ophthalmic and Physiological Optics* 27, 5 (2007), 417–439.

10. G.E. Favalora. 2005. Volumetric 3D Displays and Application Infrastructure. *IEEE Computer* 38, 8 (2005), 37–44.

11. R. Gil-Cazorla, S. Shah, and S. A. Naroo. 2015. A review of the surgical options for the correction of presbyopia. *British Journal of Ophthalmology* (2015).

12. E. Harb, F. Thorn, and D. Troilo. 2006. Characteristics of accommodative behavior during sustained reading in emmetropes and myopes. *Vision Research* 46, 16 (2006), 2581–92.

13. M. G. Harris, J. E. Sheedy, and C. M. Gan. 1992. Vision and Task Performance with Monovision and Diffractive Bifocal Contact Lenses. *Optometry and Vision Science* 69, 8 (1992), 609–14.

14. R. T. Held, E. A. Cooper, and M. S. Banks. 2012. Blur and Disparity Are Complementary Cues to Depth. *Current Biology* 22, 5 (2012), 426 – 431.

15. S. Hillaire, A. Lecuyer, R. Cozot, and G. Casiez. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In *Proc. IEEE VR*. 47–50.

16. D. M. Hoffman and M. S. Banks. 2010. Focus information is used to interpret binocular images. *Journal of Vision* 10, 5 (2010), 13.

17. X. Hu and H. Hua. 2014. Design and Assessment of a Depth-Fused Multi-Focal-Plane Display Prototype. *Journal of Display Technology* 10, 4 (2014), 308–316.

18. F.C. Huang, K. Chen, and G. Wetzstein. 2015. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-Eye Light Field Display with Focus Cues. *ACM Trans. Graph. (SIGGRAPH)* 34, 4 (2015).

19. P. V. Johnson, J. A. Q. Parnell, J. Kim, C. D. Saunter, G. D. Love, and M. S. Banks. 2015. Dynamic lens and monovision 3D displays to improve viewer comfort. *ArXiv e-prints 1512.09163* (2015).

20. A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. 2007. Rendering for an interactive 360° light field display. *ACM Trans. Graph. (SIGGRAPH)* 26, 3 (2007).

21. M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn. 2009. Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology* 53, 3 (2009).

22. D. Lanman, M. Hirsch, Y. Kim, and R. Raskar. 2010. Content-Adaptive Parallax Barriers: Optimizing Dual-Layer 3D Displays using Low-Rank Light Field Factorization. *ACM Trans. Graph. (SIGGRAPH Asia)* 29, 6 (2010), 163:1–163:10.

23. D. Lanman and D. Luebke. 2013. Near-eye Light Field Displays. *ACM Trans. Graph. (SIGGRAPH Asia)* 32, 6 (2013), 220:1–220:10.

24. S. Liu, D. Cheng, and H. Hua. 2008. An optical see-through head mounted display with addressable focal planes. In *Proc. ISMAR*. 33–42.

25. P. Llull, N. Bedard, W. Wu, I. Tosic, K. Berkner, and N. Balram. 2015. Design and optimization of a near-eye multifocal display system for augmented reality. In *OSA Imaging and Applied Optics*.

26. G.D. Love and M.S. Banks. 2014. Stereoscopic Image Generation with Asymmetric Level of Sharpness, Patent Application WO2014199127A1. (2014).

27. G. D. Love, D. M. Hoffman, P. J. Hands, J. Gao, A. K. Kirby, and M. S. Banks. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express* 17, 18 (2009), 15716–25.

28. G. Maiello, M. Chessa, F. Solari, and P. J. Bex. 2014. Simulated disparity and peripheral blur interact during binocular fusion. *Journal of Vision* 14, 8 (2014).

29. A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs. 2013. Focus 3D: Compressive Accommodation Display. *ACM Trans. Graphics* 32, 5 (2013), 153:1–153:13.

30. L. Marran and C. Schor. 1997. Multiaccommodative stimuli in VR systems: problems and solutions. *Human Factors* 39, 3 (1997), 382–388.

31. M. Mauderer, S. Conte, M. A. Nacenta, and D. Vishwanath. 2014. Depth Perception with Gaze-contingent Depth of Field. *ACM SIGCHI* (2014).

32. R. Narain, R. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien. 2015. Optimal Presentation of Imagery with Focus Cues on Multi-Plane Displays. *ACM Trans. Graph. (SIGGRAPH)* 34, 4 (2015).

33. E. Peli. 1999. Optometric and perceptual issues with head-mounted displays. In *Visual Instrumentation: Optical Design & Engineering Principles*. McGraw-Hill.

34. J. Rolland, M. Krueger, and A. Goon. 2000. Multifocal planes head-mounted displays. *Applied Optics* 39, 19 (2000), 3209–3215.

35. J. Rolland and L. Vaissie. 2002. Head mounted display with eyetracking capability. U.S. Patent 6,433,760B1. (2002).

36. B. Schowengerdt and E. Seibel. 2006. True 3-D scanned voxel displays using single or multiple light sources. *J. SID* 14, 2 (2006), 135–143.

37. T. Shibata, J. Kim, D.M. Hoffman, and M.S. Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision* 11, 8 (2011), 11.

38. S. Siegel. 1956. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill.

39. T. Sugihara and T. Miyasato. 1998. 32.4: A Lightweight 3-D HMD with Accommodative Compensation. *SID Digest* 29, 1 (1998), 927–930.

40. I. E. Sutherland. 1968. A Head-mounted Three Dimensional Display. In *Proc. Fall Joint Computer Conference*. 757–764.

41. M. Waldkirch, P. Lukowicz, and G. Tröster. 2004. Multiple imaging technique for extending depth of focus in retinal displays. *Optics Express* 12, 25 (2004).

42. S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks. 2005. Focus cues affect perceived depth. *Journal of Vision* 5, 7 (2005), 834–862.

43. G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. 2012. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (SIGGRAPH)* 31, 4 (2012), 1–11.

44. T. Zhou, J. X. Chen, and M. Pullen. 2007. Accurate Depth of Field Simulation in Real Time. *Computer Graphics Forum* 26, 1 (2007), 15–23.