

Reading text in an immersive head-mounted display

An investigation into displaying desktop interfaces in a 3D virtual environment

Cameron Grout
University of Waikato
Hamilton, New Zealand
cgrout@waikato.ac.nz

William Rogers
University of Waikato
Hamilton, New Zealand
b.rogers@waikato.ac.nz

Mark Apperley
University of Waikato
Hamilton, New Zealand
mapperle@waikato.ac.nz

Steve Jones
University of Waikato
Hamilton, New Zealand
stevej@waikato.ac.nz

ABSTRACT

This paper describes an experiment conducted as part of a larger project investigating the possibilities of using a virtual environment for users performing day-to-day computing tasks. The experiment is a user study analyzing the performance of reading tasks typical of a general purpose computing environment conducted in immersive virtual reality headsets. Results of this study are evaluated, and suggest that reading tasks can be performed with near equivalent performance in the virtual environment when compared to performance values obtained from baseline tasks on a traditional display.

CCS Concepts

•**Computing methodologies** → **Virtual reality**; *Perception*; •**Human-centered computing** → *User studies*; User interface design; •**Applied computing** → Personal computers and PC applications;

Keywords

Virtual reality; Fonts hinting; Rendering techniques; Readability; Application redirection

1. INTRODUCTION

In 2003, Czerwinski *et al.*[6] identified a significant performance advantage in using multiple monitors or large higher resolution displays. Through analyzing 15 participants performing tasks on large displays, they noted that participants experienced a 9% decrease in required time to complete a task when using a large display compared to the time required on a smaller display. Czerwinski noted that 14 of the 15 participants preferred the large display over smaller ones, reporting that it was easier to manage multiple windows and to switch between tasks on the larger display.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. *CHINZ 2015, September 3–4, 2015, Hamilton, New Zealand* Copyright is held by the owner/author(s).

Publication rights licensed to ACM.

ACM 978-1-4503-3670-3/15/09...\$15.00

DOI: <http://dx.doi.org/10.1145/2808047.2808055>

In 2013, the primary researcher conducted an on-site observation study with a medium sized company, where 15 participants were observed at hourly intervals to identify the number of applications being managed, and the organization of these windows on their screens. On average, 7 open windows were present on participants computers, with as many as 15 windows open in extreme cases. The dominant presentation method for these windows was a full-screen configuration, with free-form tiling and ‘half-maximized’ window arrangements being less favored methods of window presentation. While the majority of participants used multiple displays, 53% of participants indicated that despite their large display areas, they had difficulty identifying windows pertinent to their tasks, and indicated that window management and identification resources provided by their window managers were insufficient for facilitating quick task switching. When polled, 67% of participants indicated that following their last display upgrade (encompassing larger, or more displays) they experienced an increase in productivity, and believed that should their displays be upgraded again, they would experience a similar productivity boost.

An issue with traditional windowing systems and displays is that the amount of usable desktop space is finite, limited by the size and number of displays used. A number of factors limit how large this space can grow, including the price of larger displays, and hardware limitations preventing large numbers of displays from being connected to a computer. In order to help users utilize their limited space efficiently, many window managers provide facilities for organizing windows, through tiling methods, or options like ‘half-maximizing’ windows to occupy half of the display, and provide a number of assistive measures for identifying windows. In addition to these window management techniques, some systems provide virtual desktop options, allowing users to have multiple virtual desktops. Each of these virtual desktops display a set of windows, and can be swapped onto the physical device for display using key combinations. This allows users to partition work tasks onto different virtual spaces, organizing tasks in an efficient manner, but introduces additional cognitive load in recalling which virtual desktop contains the desired information, and requires additional action to switch desktops.

As it is clear that users could benefit from more usable screen space, alternative solutions have been proposed that utilize head-mounted displays (HMD) in order to expand

the users' working space by presenting a contiguous virtual world. In 1983, Fisher *et al.*[7] presented an immersive virtual environment (IVE) that could present virtual windows around the user in a 3D environment through use of binocular virtual reality HMDs. The interface was proposed as a general purpose system, allowing visualization of scenes and telepresence in addition to displaying information on virtual windows. Interaction with this system was accomplished through use of an instrumented glove, as well as speech and gesture recognition allowing interaction with virtual controls.

In 1993, Reichlen[10] demonstrated Sparchair, a system using a custom built monochromatic monocular HMD that presents familiar applications on a large X-windows frame buffer, with the position of a viewport into this buffer determined by orientation of the users' head. Reichlen reported that users were attracted by the facility to rapidly switch task contexts in this system, particularly for small quick tasks, and it was noted that there was a distinct advantage in having an "unlimited number of windows open at once"[10], reducing mouse movements and clicks normally associated with task switching on a traditional display.

Lawton[8] identified in their 2006 article that the primary barrier to common adoption of virtual reality technologies was the lack of affordable VR devices with specifications that would make them desirable. In recent years, announcements for consumer priced virtual reality hardware with competitive specifications by Oculus, Sony, Valve and others has driven consumer interest in such devices. While the initial target market for these devices is identified as the gaming and entertainment industries, companies such as Bloomberg have begun investigating the use of these devices in areas such as finance[2]. With the release of these affordable devices, it is expected that it will begin to be commonplace to see them in homes and businesses. It is also expected that consumers will identify that through the use of such devices, digital workspaces can be expanded, which will drive demand for general purpose interfaces and environments that provide access to traditional applications using these devices.

As a first step towards building a system, we needed to identify common tasks performed by users. In the preliminary study discussed above, we observed that users were employing large numbers of windows to make a large amount of reference material available for their work. In most cases this material was textual in a flow or a tabular format.

More generally, in a diary study conducted in 2010 by White *et al.*[13] roughly 400 Americans aged 20 years and older reported "all reading, writing and computer activities"[13, p. 285] for one working and one non-working day within a week. Analysis of this data by White revealed that on average participants spent 312 minutes on working days, and 244 minutes on non-working days engaging in reading tasks taking longer than 1 minute.

Clearly in order to support general computing tasks, an interface must facilitate reading text with ease and comfort. To address this, we undertook a study of reading in an immersive virtual environment. Our goal was to determine the best way of displaying textual information.

This paper describes a user study analyzing the performance of reading tasks inside a general purpose computing environment for use within virtual reality headsets. Section 2 describes the proposed system; Section 3 & 4 describes our

user study and presents the results; Sections 5 & 6 present our conclusions.

2. SPATIAL INFORMATION DISPLAY

In our work, we are building a body-stabilized virtual environment (view is "fixed relative to the users body position and varies as the user changes viewpoint orientation"[4]). The system is designed for use with the Oculus Rift series of immersive virtual reality displays, that is intended to be used in a seated configuration. This environment is written in C#, using Win32 APIs via Platform Invocation services, and DirectX through use of the SharpDX library. This environment gives users the impression of being surrounded by a virtual sphere, upon which application windows can be placed and freely moved (See Figure 1). Application interfaces are captured and interacted with through use of application redirection techniques discussed by van Dantzich *et al.*[12], though the implemented system is significantly cruder than the reference system due to the lack of access to required source code.

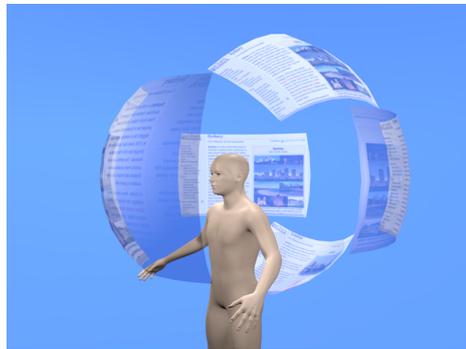


Figure 1: Visualization of the developed environment, showing curved panels

Navigation inside this environment is accomplished through head movement, detected by sensors provided with the Oculus devices. Orientation is detected via three degrees of freedom (3doF) sensors onboard each headset, which provide data collected from gyroscopic, magnetometer and accelerometer sensors at 1000Hz. Orientation data obtained from these sensors is used to modify the viewport that is seen by the user, by rotating the virtual cameras to match the real-world head rotation. The DK2 additionally provides positional data through use of an external infrared camera, which detects the location of the headset within a $72^\circ W \times 52^\circ H$ frustum at a range of 0.5m–2.5m from the camera. When using the DK2 headset, this positional data is used to modify the position of the virtual camera in concert with real world movements, allowing the user to move their eyes closer to an item in order to see it better amongst other uses. It is our intention that interaction in the environment will be performed using a traditional mouse and keyboard interface. This will preserve familiar interactions, such as click-dragging on an application window decoration in order to move it, and resizing applications by grabbing the border or handle located in the bottom-right of the window. Applications window positioning will obey z-ordering rules, allowing windows to overlap one-another, as in traditional desktop environments.

An unknown at this point in development is how well users will be able to read text. This is vital because general computer users spend a large portion of their time reading (as noted in Section 1). Issues include text size, window placement and window shape. In order to investigate this, a preliminary prototype of this system was built that displays static images rather than live applications. This prototype implemented all the navigation and display features, and a functioning but incomplete interaction system. Items displayed in this system can be mapped onto planes, or spherically projected curved panels of different sizes.

3. USER STUDY

Using the prototype system, we were able to study questions that our early experiments raised concerning reading in a virtual environment. A number of these questions were addressed using reading performance metrics, comparing accurate reading speeds across devices in order to provide a reasonable indication of performance for the study. A reduction in reading performance would indicate a likely reduction in productivity when compared to the participant’s normal performance, and vice versa.

During informal testing there had been no consensus as to which headset was preferable, despite their numerous differences (detailed in Table 1). Because the DK2 has a higher resolution, initial expectations were that, in spite of its reduced field of view (FOV) it would be preferred over the DK1, but general feedback given was that they were both equally suited for interacting with the environment. In order to determine if a performance difference was present, it was deemed necessary to perform experiments on both devices, and to compare the results.

Table 1: Oculus Rift Version Features

	DK1	DK2
Resolution (pixels/eye)	640x800	960x1080
FOV (Nominal)	110°	100°
Tracking	3doF	6doF
Pixel Layout	RGB	RGBG PenTile
Weight	380g	440g

During these informal tests, a number of users indicated that they felt that curved applications were harder to read because they felt unnatural. These users expressed a preference for flat panels, indicating that these felt more like reading information from a screen – something they were very familiar with – and it was believed that these flat panels allowed for faster reading due to this feeling of familiarity. This sentiment was not shared by all users however, so this indicated that an investigation into these distortion options, and their effect on performance, was desirable.

For informal testing, the distance between all applications and the environment origin had been fixed to a value determined to be suitable by the primary researcher. As this position was chosen arbitrarily, it is possible that users were required to physically move to get themselves in a position in the virtual environment where they could read comfortably. In order to provide a default position, preferred distances should be collected, allowing a well reasoned average position to be determined.

3.1 Approach

In order to address these questions and concerns, an experiment was designed utilizing a modified version of the developed environment. In place of the normal live applications, static images of an application were displayed, and the user given the ability to change the distance at which the image was placed relative to themselves. Users would then read textual content from the application image, and the application position, reading speed and reading accuracy would be recorded.

In order to ensure the size of the application window was not a confounding factor, multiple sizes of page were used, with one of each size being displayed as a flat panel, and one as a curved panel. Each participant was tested with both headsets and was shown the same number of samples in each. In order to obtain a basis for comparison, users were also presented with images of the same size displayed on a traditional computer screen, and the same statistics were recorded.

3.2 Hypotheses

The study set out to test several hypotheses, aiming to address each of the questions and concerns raised. The hypotheses were as follows:

1. Curved display panels will be preferred over flat panels for all display sizes
2. An increase in reading performance will be noted in all experiments using a curved panel when compared to that of a equivalently sized flat panel
3. Study participants will prefer the traditional computer monitor over either of the headsets, and will prefer the DK2 over the DK1
4. Reading performance will peak using the traditional monitor, and will be followed by the DK2, then finally DK1
5. Participants will identify an optimal reading distance in the virtual environment, and we will observe a near constant placement at this point for all participants

3.3 Participants

Participants for the user study were volunteers sourced from the university student population, with the majority being from the Faculty of Computing and Mathematical Sciences. 24 participants were selected from the pool of volunteers with visual acuity the deciding factor in selection. Of the participants, 16 were male and 8 were female; 2 of these participants had their sight corrected through the use of contact lenses, and 3 had congenital issues affecting a single eye that could not be corrected. This is visualized in Figure 2.

Subjects self-reported their preferences for reading content, with 23 participants indicating a strong preference for reading on-screen, and 1 participant indicating a strong preference for reading physical paper-based media. Of the participants, only 1 had extensive experience with virtual reality devices, with the remainder indicating very slight, or no experience with VR.

3.4 Equipment

Two immersive virtual reality headsets, the Oculus Rift DK1 (DK1) and the Oculus Rift DK2 (DK2), were used as the headset devices for this experiment. Both headsets

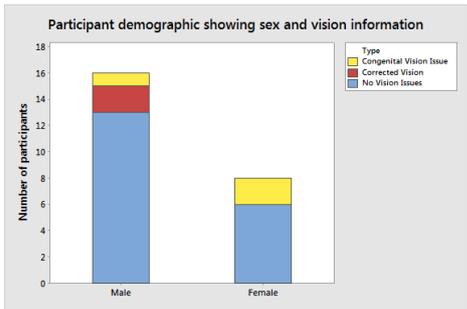


Figure 2: Study participant demographic information showing numbers of participants with vision issues

supply multiple lens sets which can be swapped as required by the user. Both headsets were fitted with the ‘A’ lens sets, intended for use when no vision issues are present, and the eye relief dials were set to their maximum extent. The DK1, which includes configurable hardware brightness and contrast setting, was configured such that these levels were set the middle of their operational levels, which gave the display a similar appearance to a traditional display. As the DK1 does not provide positional data, the positional tracking feature of the DK2 was disabled.

A Dell U2312HMi LCD monitor was used as the traditional display. This display contains a 23” panel, with a 60Hz refresh rate and native resolution of 1920x1080. The display was reset to factory settings prior to commencement of the study, with brightness and contrast set to 75% of their maximum values.

The testing computer was a Dell Optiplex 9020 with an quad-core Intel i7 4770 clocked at 3.40 GHz, with 16GB of DDR3 RAM and an NVidia GeForce GTX 750 Ti. The operating system used was Windows 7 SP1 64-bit, updated with the latest hotfixes and updates available at the time of testing. Peripherals used were a traditional 2-button mouse with scroll-wheel, and a 104-key US-layout keyboard.

The software running the experiment was the prototype system described in Section 2. Interaction was managed through the keyboard arrow keys, to advance or retreat through samples, and the mouse scroll wheel, to move the panel toward or away from the participant. When the sample was changed, the sample’s position in the environment was set to an unreadable distance in order to force the participants to position each panel independently. This application displayed output on the headsets used, as well as on a secondary monitor observed by the experiment supervisor, and operated in the 60-75 FPS range.

3.5 Experiment and Method

Fifteen page samples were selected from the Simple Wikipedia collection of major world cities. From each sample, a paragraph was selected that consisted of 110-130 words, with a Flesch–Kincaid¹ reading score of between 60 & 65 (as measured by Microsoft Word, 2013). Each of the selected sample

¹A readability score derived from the Flesch readability formula, defined as $RE = 206.835 - (0.846 \times NS) - (1.015 \times \frac{W}{S})$, “Where [NS] is the average number of syllables per 100 words, and $\frac{W}{S}$ is the average number of words per sentence”[3].

paragraphs was modified to remove superfluous visual noise, such as inline reference marks, and to remove words which would be difficult for someone unfamiliar with them to pronounce. 5 samples were then displayed at each of the resolutions specified in Table 2 using Google Chrome running on Windows 7 with factory default settings, then a screenshot of the sample was taken. The collection of sample images were then split into 3 sequences as specified in Table 3.

Table 2: Sample sizes and aliases

Resolution	Aspect Ratio	Name
800x600	4:3	Small
1366x768	16:10	Medium
1920x1080	16:9	Large

Table 3: Sample collections

Sequence #	S	M	L
Seq 0	2	2	2
Seq 1	1	1	1
Seq 2	2	2	2

Each participant was tested on two headsets, and one traditional computer screen. The order of these tests was assigned in an alternating fashion to participants as DK1, Monitor, DK2 or DK2, Monitor, DK1. The baseline test using the monitor was always the second device, as it gave the participant’s eyes a chance to rest in between headset runs. A sequence of sample images was assigned to each headset device in an alternating fashion, with the monitor device always being assigned Seq 1. Within each sequence, the order of display was randomized between participants and devices. For each sequence displayed on a headset, one of each sample size was displayed using a curved distortion, and the other with no distortion (flat).

For experimental runs using the traditional computer screen, participants were seated facing the display, and positioned such that their eyes were approximately 750mm away from the display. The height and angle of the display were then adjusted to ensure that the participant’s line of sight was perpendicular with the center of the display. For experimental runs using a headset, the participant was fitted with the appropriate device such that it was fixed firmly but comfortably to their head. They were then seated in front of a desk, and provided with a keyboard and mouse. A sample image was then displayed centered on the screen with black borders filling in any additional space.

For experiment runs using headsets, participants were presented with a sample image in the environment, loaded at an unreadable distance. The participant was then prompted to move the panel using the mouse scroll wheel until the sample was at the far edge of a comfortable reading distance. After positioning the panel, the panel position was logged and the participant was prompted to release the controls.

Following the setup for each experimental run, participants were shown each sample in sequence. For each displayed sample, the participant was prompted to begin reading the sample at a specified point in the text until asked to stop. As the participant read, the range and pattern of their head movements were recorded by the experiment supervisor, and their responses were recorded using audio recording devices. This procedure was repeated until all applicable samples had been read, then a questionnaire was provided for the participant to fill in.

These questionnaires asked the participants if they encountered any difficulties reading the sample images, and if so, were prompted to indicate their perceived cause of the difficulty. For experiment runs using the headsets, participants were also asked if they had a preferred display method (curved, or flat), to indicate if they felt comfortable using the devices, and to rate any discomfort they had felt. Following the completion of all 3 experiment runs, the participants were asked to fill in one additional questionnaire, which asked them to rank the experimental devices in order of preference, and give reasoning for their decision.

3.6 Data Treatment & Analysis

Following the experiment, recorded participant responses were edited to remove comments to the experimental supervisor, as well as general comments. These included, but were not limited to, comments like ‘this is an odd sensation’ regarding reading in the headset, ‘I didn’t know that’ when reading novel information. Comments that were not deliberate asides and were a part of the thinking process such as ‘I think’ after reading a word, and filler words and noises like ‘Umm’ and ‘Ah’, were not removed from the recordings. Any removals from the recordings were performed against the edges of the waveform that comprised the comment for removal. Any surrounding silence or words were left in place. The length of time required to read each sample was recorded from the point in the recording where the participant began to speak, until the end of the last word of the sample.

These modified recordings were then analyzed using running record techniques as described by Clay[5] to determine the accuracy of the reading. With this analysis, errors, self corrections and repetitions were recorded based on the rules described for the method. An error was recorded for each missed, substituted or inserted word. Neither contractions or initialisms were penalized, allowing ‘it’s’ and ‘it is’ substitutions, as well as ‘US’ and ‘United States’ expansions or contractions. Proper noun errors were excluded from the error count, so long as the error could be attributed to unfamiliarity with the word (eg, ‘Chesapeake’ said as ‘Cheese Peak’). If a participant made an error, then corrected themselves, the error was not counted, but was instead recorded as a self correction.

Accuracy of reading was calculated for each participant and sample using the word count of the sample, and the number of errors made by the participant while reading that sample using the equation:

$$AR_{sample} = 100 - \frac{errors}{wordcount} \times 100$$

In addition to calculating the accuracy of reading, an *Accurate Words per Minute* (AWPM) value was calculated with the wordcount of the sample, the number of errors made while reading the sample, and the time required to read the sample (TTR). This was evaluated as:

$$AWPM_{sample} = \frac{wordcount - errors}{TTR_{seconds}} \times 60$$

Following the calculations for AWPM for all participants and samples, the AWPM of each sample read while using a headset was compared against that participants own baseline AWPM calculated from their sample of the same size read from the traditional monitor. This was accomplished

with the equation:

$$AWPM_{diff} = \frac{AWPM_{baseline} - AWPM_{sample}}{AWPM_{baseline}}$$

Information written by the testing application to log files was processed to determine the horizontal angle subtended for a lowercase letter ‘m’ for each sample at the location chosen by the participant. The letter ‘m’ was chosen for evaluation, as it was indicated as one of the more difficult characters to identify during headset experiments. Including color fringing caused by font-hinting, the lower case ‘m’ in the sample images occupied 12×8 pixels before being displayed in the headsets. The subtended angle was calculated by determining the angle between two 3-D vectors positioned at the center-left and center-right of the character bounding box. For flat panels, both best-case and worst-case angles were calculated, while for curved panels only a single value was needed as the nature of the curve distortion provides a constant angular character pitch across its entire area. Best case samples, as well as the one sample for curved samples, were taken with the assumption that the character to be measured was centered at the $(0, 0, +Z)$ position vector, where Z is the in-engine distance the panel was positioned at by the participant. Worst-case calculations for the flat samples were taken with the character to be measured positioned at the top-left extreme of the sample area. The character positions as measured on a flat panel can be visualized in Figure 3.

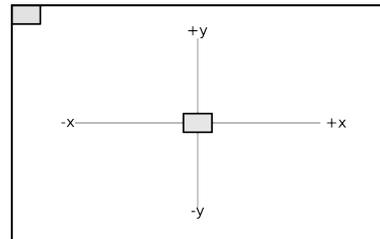


Figure 3: Locations of character samples used for calculating horizontal angle subtended relative to the sample in question. The box in the top-left shows the worst-case measurement, while the box in the center shows the best-case

4. RESULTS

From the final questionnaire completed at the end of the user study, the device preferences indicated by participants were collected, and presented in Table 4. From this data, a clear pattern can be seen, with the traditional monitor being indicated as the most preferred device by 92% of participants. 75% of participants indicated that the DK2 was their second choice of device, and finally 79% indicated that the DK1 was their least preferred. From the feedback given in this questionnaire it was found that most participants preferred the traditional monitor because of its familiarity and clarity, indicating that the text was crisp and easy to read, and that the amount of practice they had at reading from these types of displays meant there was nothing they had to grow accustomed to in order to complete the reading task. A minority of participants indicated a preference for the DK1 over the DK2, indicating that while reading was

more difficult using the DK1 as the amount of head movements required was large, the clarity of text inside the DK1 was better than that of the DK2. With reference to hypothesis #3, we can conclude that the hypothesis is true, in that the greater proportion of the participant set indicated the device preference order hypothesized.

Table 4: Overall device preferences

Device	1 st	2 nd	3 rd
Monitor	22	2	0
DK2	1	18	5
DK1	1	4	19

Table 5: Panel distortion preferences

Distortion	DK1	DK2
Flat	2	2
Curved	16	19
No Pref.	5	3

From evaluating the responses given in questionnaires following experiment runs on each device, the preferred distortions indicated by participants for each headset were collected, and are shown in Table 5. As can be seen from the table, the curved distortion method was most preferred amongst the participants, with 67% indicating it as their preferred method on the DK1 headset, and 79% indicating the same on the DK2. 2 participants indicated that they preferred flat panels on both headsets, indicating that while the curved panels were easier to read when the samples were large, the curve felt unnatural and was likened to ‘reading off the side of a building’. All participants indicated that when reading larger samples the curved distortion made reading easier, but noted that with smaller samples there was no notable difference in reading ease between curved and flat panels. With respect to hypothesis #1, we can conclude that for large samples, the curved distortion was preferred by the majority of participants on both headsets, but there was ambivalence about the distortion method with smaller samples.

Calculated differential AWPM values were evaluated using a paired t-test over the results for flat and curved samples of the same size. This information is displayed in Table 6, and is shown graphically in Figure 4. From examining the mean difference values between samples, it can be seen that for the small and medium samples on the DK2 and the small samples on the DK1 that the difference in performance for these samples were not statistically significantly different ($p \gg 0.05$). However for the large samples read on both headsets as well as the medium samples read on the DK1, large mean difference values are present, indicating that there was significant performance differences ($p = 0.000$) between the samples when they were presented in a flat or curved manner. This allows us to address hypothesis #2, and state that while the hypothesis holds true for large samples, it does not hold for smaller ones.

Referring to Figure 4, it can be seen that when comparing samples of the same size and distortion method across devices, the DK2 consistently has a smaller percentage difference from the baseline when compared to the performance on the DK1. With reference to the data contained in Table 6, it can be seen that small samples on both devices and either distortion method fall within standard error of

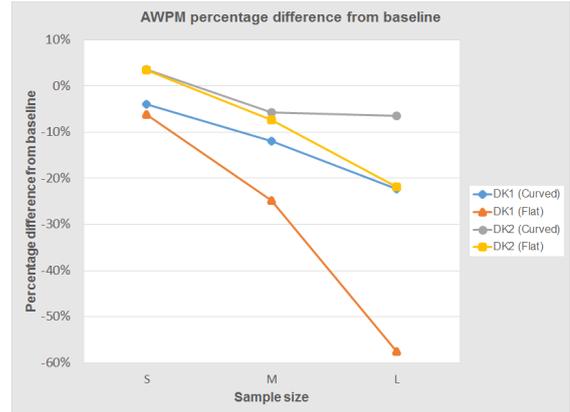


Figure 4: Mean AWPM percentage differences for all sample sizes and devices

the baseline value, indicating that for small samples, both devices are near equivalent in performance to the baseline. This is not the case however with medium samples, where the flat panel read on the DK1 plummets below the baseline performance. The remaining medium samples fall within standard error of the baseline performance, with the samples read on the DK2 having means lying closer to the baseline and with smaller standard error values than that of the DK1 curved sample. Large samples present a large spread of mean percentage differences, with the large curved sample read on the DK2 still falling near the baseline within standard error, and the remaining samples falling below the baseline completely. With respect to hypotheses #4, we can conclude that for small samples, there is no statistically significant performance difference across the 2 headsets and traditional monitor. However, as the size of samples increases, the hypothesized pattern begins to emerge with all devices falling below the baseline in combined performance across distortions.

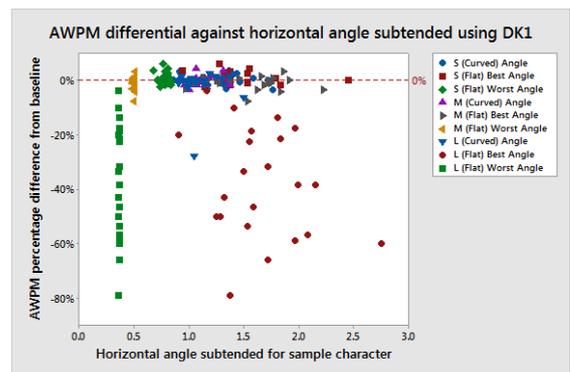


Figure 5: Horizontal angle subtended for a lower-case letter ‘m’ against participant running record accuracy compared to baseline for samples using the DK1 headset

For each participant and sample, the best and worst case subtended angles were calculated and compared against the differential reading accuracy of the sample from the baseline. This information can be visualized for the DK1 experiments in Figure 5, and for the DK2 experiments in Figure 7.

Table 6: Paired T-tests for distortions pairs, over the percentage difference from baseline of same sample size

Device	Size	$\mu(\text{Flat})$	$\mu(\text{Curve})$	$\mu(\text{Difference})$	95% CI	t	p
DK1	L	-57.59% \pm 14.32	-22.25% \pm 17.36	-35.33%	-43.15% to -27.52%	t(21) = -9.41	0.000
DK1	M	-24.87% \pm 13.66	-12.00% \pm 15.43	-12.87%	-18.45% to -7.30%	t(22) = -4.79	0.000
DK1	S	-6.12% \pm 15.20	-3.90% \pm 20.36	-2.22%	-9.50% to 5.05%	t(22) = -0.63	0.533
DK2	L	-21.88% \pm 14.58	-6.51% \pm 9.63	-15.37%	-21.93% to -8.82%	t(23) = -4.85	0.000
DK2	M	-7.39% \pm 10.36	-5.70% \pm 7.76	-1.69%	-6.25% to -2.86%	t(22) = -0.77	0.449
DK2	S	3.51% \pm 11.85	3.57% \pm 16.89	-0.06%	-7.12% to 7.00%	t(23) = -0.02	0.987

A broad cluster of points can be seen in each graph centered around the 0% differential accuracy line, showing that while there was little difference in reading accuracy for most samples, there was notable variation in placements of samples inside the environment. For the best case values only, this cluster represents a mean best case value of $1.35^\circ \pm 0.33$ for the DK1 samples, and $0.97^\circ \pm 0.26$ for the DK2.

For the large flat panels on the DK1, it can be seen that although the best case angles are spread over a wide range and are larger than the best case mean, the differential accuracy decreased. This effect can be explained by the narrow worst case angles that accompany the best case options, which shows that while there was a large degree of angular variation in the best case, in the worst case the subtended angle is near constant. The flat medium panel worst case markers can be seen to the right of the large markers, and also show a near constant subtended angle, but with a greatly reduced differential accuracy score. This indicates that as the worst case angle gets narrower, the overall performance of the sample will decrease. This effect can be seen in the DK2 results also, with the large flat panel showing a narrow angle, with reduced accuracy, however the differential accuracy values are notably better than those expressed for the DK1 sample. This difference in differential accuracies despite the same sized samples occupying near identical angles across devices, suggests that the device specifications affect the reading performance, likely through differences in pixel density.

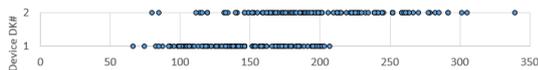


Figure 6: Calculated distance (mm) from the participant to the chosen panel position

With regards to hypothesis #5, we can conclude that while participants did not identify a single optimal location for placement of panels, they did identify a range of locations that would be suitable (see Figure: 6). Large flat panels presented as outliers with significant variance in placement, and an equally notable variation in accuracy differentials, suggesting that the size of these panels exceeds what would be considered reasonable to display at the headset resolution with no panel distortion.

5. EVALUATION

Over the duration of the user study, a number of issues were indicated by participants that fell outside of the scope the study was intending to address. The most prominent of these was a perceived ‘fuzziness’ of text, predominantly experienced on samples in the DK2. The most common description of this problem was seeing a shadow of color sur-

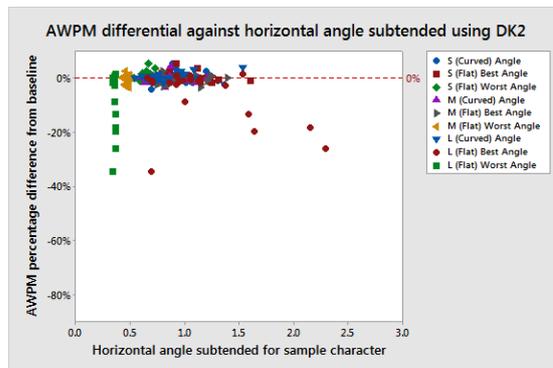


Figure 7: Horizontal angle subtended for a lower-case letter ‘m’ against participant running record accuracy compared to baseline for samples using the DK2 headset

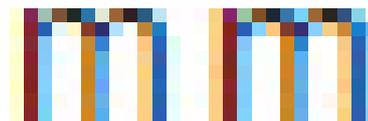


Figure 8: Letter ‘m’ extracted from a sample image (left), zoomed to display the ClearType color fringing. The same ‘m’ with faint colors amplified is displayed on the right

rounding the text, causing the characters to be difficult to read because of the increased graphical noise. The cause of this on the DK2 appears to relate to Microsoft’s ClearType, which applies subpixel rendering techniques to text in order to improve the look of text on LCD RGB displays[1]. As the samples that were displayed were taken as screenshot images from a system utilizing ClearType, fonts were decorated with color fringing (see Figure 8) that would not normally be readily noticeable on an LCD display due to the known arrangement of subpixels that the ClearType engine targets. The display in the DK2 however, is an RGBG PenTile display with subpixels rendered in a diamond pattern. As this layout is different from that which the ClearType engine is expecting to be rendering to, the effects are not what was intended, and a blurriness is observed.

In addition to the color shadowing issue observed, some participants noted that they had difficulty reading when their head was stationary, but no problems reading when moving their heads. As the image samples being read were taken at a high resolution, these had to be scaled down to display on a smaller number of physical pixels inside the headset. This texture filtering process produces interesting results in conjunction with font-hinting and the kerning

of certain characters, producing words that cannot be easily read. However, when moving the headset, the view is constantly being redrawn, and the texture filtering effects recalculated, leading to differing images being produced on each frame. This can be seen in Figure 9, where the word ‘while’ is seen through the lenses of the DK2 over 9 consecutive frames. Issues can clearly be seen with displaying a white pixel column between the ‘i’ and ‘l’ characters, with some frames blurring these two characters together, and the ‘h’ and ‘e’ characters appear with many of their defining features obscured on some frames.



Figure 9: The word ‘while’ as seen over 9 frames through the lenses of the DK2

In order to address these issues in future experiments, trials with different fonts and the use of ClearType will be conducted. Sheedy *et al.*[11] note that while the effects of ClearType font hinting is preferred by readers, there is no improvement in text legibility, reading speed or comfort with it turned on or off. Due to this, it is likely that future experiments will be conducted with it off in order to remove a confounding factor.

Additional issues noted by some participants were varying degrees of double vision and chromic aberration, which were alleviated by adjusting the Inter-Pupillary Distance (IPD) value. These IPD values were calculated using a utility that is a part of the Oculus Runtime, but this tool appears to suffer from imprecision, which becomes an issue when using the devices. McCleary[9] notes that this double vision ‘prism’ effect can be avoided with accurate measurements of the IPD using a digital pupilometer, and for future experiments we will attempt to utilize one of these in order to avoid these issues.

6. CONCLUSIONS & FUTURE WORK

Our results show that it is possible for users to perform traditional reading tasks inside a immersive virtual environment with near-baseline performance under ideal circumstances. We identified that rendering application windows onto flat panels significantly impaired reading performance on larger panels, due to the effects of individual characters occupying a narrower subtended angle of vision when rendered near the edge of the panel when compared to characters displayed at the center. This, coupled with the discovery that curved distortion had no notable effect on performance when compared to small flat panels, suggests that future work should focus on utilizing curved panels exclusively. It was found that there were significant performance differences when comparing the results for either headset, with the DK2 performing better than the same sample on

the DK1, which when combined with the positional tracking functionality the DK2 provides over the DK1, allows us to pursue the DK2 as our primary development target.

As development of the testing application continues, information gleaned from this study will direct the development, and will allow us to make decisions based on quantitative data. Future investigations will be performed, investigating issues identified by this study, with an emphasis on font-hinting techniques, and their suitability for use inside these environments.

7. REFERENCES

- [1] Microsoft Research, ClearType. <http://research.microsoft.com/en-us/projects/ClearType/>.
- [2] 2014. Virtual reality headset Oculus Rift meets the Bloomberg terminal. <http://qz.com/218129>.
- [3] J. C. Alderson. *Assessing Reading*. Cambridge Language Assessment Series. Cambridge University Press, 2000.
- [4] M. Billingham, J. Bowskill, N. Dyer, and J. Morphet. An evaluation of wearable information spaces. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, pages 20–27, 1998.
- [5] M. M. Clay. *Running Records for Classroom Teachers*. 2000.
- [6] M. Czerwinski, G. Smith, T. Regan, B. Meyers, G. Robertson, and G. Starkweather. Toward Characterizing the Productivity Benefits of Very Large Displays. *Interact*, 3:pages 9–16, 2003.
- [7] S. S. Fisher, M. McGreevy, J. Humphries, and W. Robinett. Virtual environment display system. In *Proceedings of the 1986 workshop on Interactive 3D graphics, I3D '86*, pages 77–87, New York, NY, USA, 1987. ACM.
- [8] G. Lawton. Making virtual reality more accessible. *Computer*, 39(6):pages 12–15, 2006.
- [9] D. S. McCleary. *The Optician Training Manual*. Santa Rosa Publishing, Temecula, CA, 1 edition edition, Aug. 2009.
- [10] B. A. Reichlen. Sparchair: A one hundred million pixel display. In *Virtual Reality Annual International Symposium*, volume 0, pages 300–307, Los Alamitos, CA, USA, 1993. IEEE Computer Society.
- [11] J. Sheedy, Y.-C. Tai, M. Subbaram, S. Gowrisankaran, and J. Hayes. ClearType sub-pixel text rendering: Preference, legibility and reading performance. *Displays*, 29(2):pages 138–151, Mar. 2008.
- [12] M. van Dantzich, V. Gorokhovskiy, and G. Robertson. Application redirection: hosting Windows applications in 3d. In *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM international conference on Information and knowledge management, NPIVM '99*, pages 87–91, New York, NY, USA, 1999. ACM.
- [13] S. White, J. Chen, and B. Forsyth. Reading-Related Literacy Activities of American Adults: Time Spent, Task Types, and Cognitive Skills Used. *Journal of Literacy Research*, 42(3):pages 276–307, Sept. 2010.