Considering the effects of display persistence on eye movements and readability in virtual reality

***Xiuyun Wu*[*,**], *T. Scott Murdison*[*]**
***Reality Labs at Meta, Redmond, Washington**
****University of British Columbia, Vancouver, Canada**

Abstract

We measured reading speed and oculomotor metrics in a novel saccadic rapid-serial-visual-presentation task to understand how persistence affects common activities in head-mounted displays. Surprisingly, we did not observe any effects of persistence on reading performance or oculomotor metrics. We discuss the implications for current and future VR and AR displays.

readability; eye movements; phantom array; temporal light artifacts

1. Introduction

A fundamental difference between natural and artificially displayed scenes is the temporal properties of illumination. In natural scenes, photons are continuously emitted and projected onto the retina. In electronic visual displays such as VR/AR head-mounted displays (HMDs), images are illuminated at a discrete frequency, defined as the refresh rate (frames per second) for a specific illumination period. This illumination period can be described in either absolute units as persistence (ms) or frame-relative units as duty cycle (percentage of frame time). These temporal display parameters constrain the accuracy to which virtual motion can be reproduced, and greatly impact the display's overall power usage. In addition, they have been shown to cause both visible artifacts[1,2] and oculomotor control deficits[2]. Therefore, it is crucial to balance the visual user experience against the temporal properties of HMDs[1,2]. To date, metrics to perform this optimization have taken the form of perceptual[1,2] and oculomotor[2] thresholds derived under controlled experimental settings. It remains unclear how user experience in more realistic HMD use cases will be affected, as performance in complicated tasks would depend on a variety of factors. Here we seek to examine how display-persistence related deficits in perception and oculomotor control would impact the visual user experience in reading, a typical HMD use case rich with eye movements (saccades).

Specifically, we focus on the effect of persistence. HMDs require low persistence to eliminate visible blurring of the retinal image during head motion and the resulting vestibulo-ocular reflex (VOR), which stabilizes gaze on world-fixed objects. However, lowering persistence comes at a cost. It is the main driver of the perceptual artifact named strobing (a.k.a. 'ghosting' or 'phantom array')[1,3] during rapid ballistic eye movements, or saccades, and has also been shown to cause saccadic targeting errors[2]. Strobing results from a lack of retinal blur during saccades due to low display persistence, and appears to the user as multiple copies of a virtual object across the visual field. Its appearance is believed to be modulated by display properties such as content sparseness, luminance, and contrast[1,3,4]. The saccadic errors associated with low persistence can be costly for discriminability of the target features[2], as visual acuity drops quickly with retinal eccentricity[5].

Even if visible strobing and saccadic targeting deficits do occur at low persistence levels, their existence does not directly speak to how strongly performance will be affected in a more realistic HMD setting. Reading is a good candidate task to test any use-case relevant persistence effects while still providing experimentally tractable results, as reading efficiency depends on both the image clarity[6] and accurate targeting of saccades[7]. On one hand, reading speed could be compromised due to strobing and saccadic deficits. The perception of strobing would cause ambiguity of the text, and the saccadic undershooting deficit[2] might hinder reading speed due to suboptimal landing positions—deviation of the initial fixation away from an optimal viewing position has been shown to result in longer gaze durations during reading[7]. Previous work[8] found this optimal viewing position to be near the word center, presumably minimizing peripheral crowding which slows down reading speed[9].

On the other hand, the perception of strobing is highly variable across observers[2]. In addition, image contrast affects saccade amplitudes in visual search, and the effect is dependent on the search target (letter vs. computer icon), potentially due to the dependence of information at different oriented spatial frequencies (SFs)[10,11]. If the saccadic deficit under low persistence is due to low fidelity of neural target representation[2], it might also be modulated by sensory image statistics (e.g. contrast, luminance, SF, structure) that alter the fidelity of that representation. Therefore, a lack of

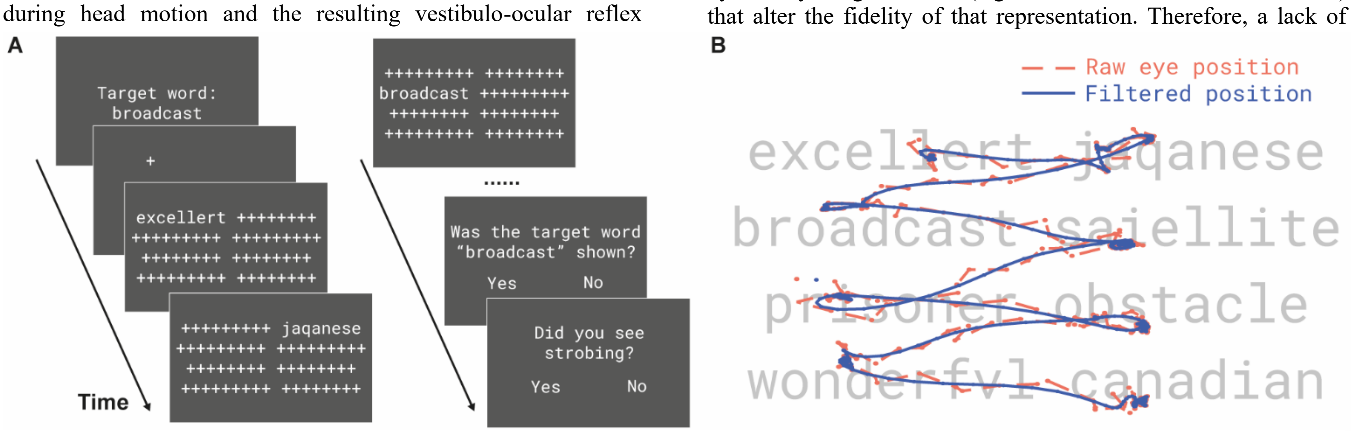

Figure 1. *(A) Example of one trial in the reading task. During the word list display, following the natural reading order in English, each word is shown sequentially for the same duration within the trial. The display duration of each word is controlled adaptively across trials. (B) Example of eye position in one trial. The red dashed line indicates the raw eye position, and the blue solid line indicates the filtered eye position.*

persistence effect on reading is also possible due to the individual variability in strobing perception, and other image-based factors.

In the current study, we use a novel saccadic rapid-serial-visual-presentation (RSVP) task (Fig. 1A, see more details in Methods) to examine the effect of persistence on silent reading speed, and quantify whether reading speed is correlated with changes in saccadic eye movement dynamics. The RSVP task without eye movements has previously been used to study readability in additive displays[12], and we adapt it to imitate a more natural reading condition with eye movements. The results are intended to inform an improved HMD persistence-optimization range based on task performance metrics.

2. Methods

2.1 Apparatus

We used a modified off-the-shelf VR HMD, Oculus Rift CV1, with refresh rate set at 90Hz. The display has an approximate per-eye field of view of 93×101 degrees of visual angle (deg), resolution of 1080×1200 pixels, with an approximate pixel density of 14 pixels per degree. We used an HMD-integrated binocular XR eye tracking platform from Tobii (Tobii AB, Sweden). This eye tracker has a sampling frequency of 240 Hz and is based on Tobii's latest generation off-axis (direct to eye) solution for VR and AR optical designs – including 'pancake' lens designs common in newer VR products. Using a custom 5 point head-fixed calibration and validation (arranged in a cross), all participants measured a 95th percentile accuracy (across all points) of <3.2 deg and 95th percentile spread of <2.3 deg (note that the calibration quality differs among participants, and these are the worst values).

2.2 Stimuli

We used word lists as the reading material. Each list consists of eight words (8-9 letters per word), randomly chosen from the top 5000 frequent words in American English. Swear words and emotionally triggering words are removed from the list by manual inspection to avoid confounds due to word saliency. To increase task difficulty while making the test more sensitive, we introduced random misspellings. Each word has a 50% chance to have one letter replaced by a similar looking letter (e.g. "c" replaced by "o"). To keep the difficulty constant, the first and last letter of the word will not be replaced as they are less affected by crowding with fewer flankers compared to other letters, thus easier to recognize[13].

We used the Roboto Mono font as it is a widely used font optimized for readability on screens under a variety of situations. More importantly, it is monospaced, thus giving the same advanced width (mark + white space) for each letter. The size of the font was 26 pt, corresponding to an advanced width of approximately 2.4 deg per letter. We tested 5 levels of persistence values, corresponding to duty cycles of 8%, 10%, 15%, 20%, and 25% at the Rift CV1's 90 Hz refresh rate. Since stimuli at higher persistence levels would emit more light due to longer illumination duration, we calibrated each color channel's digital value so that the same color would have the same luminance and chromaticity (xy in CIE color space[14]) across persistence levels using a Konica Minolta CS-160 luminance and color meter. Text stimuli were shown in white (35.34 ± 0.83 cd/m^2) on a dark grey background (0.86 ± 0.18 cd/m^2), having similar contrast (difference:average = 1.91 ± 0.02, or stimuli:background = 42.37 ± 7.64) across persistence levels.

2.3 Procedure

We recruited a total of 9 participants (age 24-52, 6 females and 2 males). Each reported to have normal or corrected-to-normal visual acuity, had no known neurological disorders, and were fluent in written English. At the beginning of the experiment, participants were briefed and shown demos of strobing, to ensure sensitivity to this artifact during eye movements.

At the beginning of each trial, a target word was shown for 1.5 s (see Fig. 1A). Then a fixation cross was shown at the location of the center of the first word for a random duration between 0.8 to 1.2 s, to prevent anticipatory saccades to the second word before it was revealed. Participants were asked to fixate at the cross, then view the word list, and report whether the target word was shown. The target word was chosen randomly from the second to the last words in the list, and only the correct spelling was considered as "shown". During the word list display, only one word was revealed at a time, while the other words covered by placeholders (crosses). In order to read each word, participant would follow the word display with saccades (Fig. 1B). Participants were instructed to finish reading the word list even if they spotted the target word in the middle of the word presentation. The reading speed (or display duration of each word) was controlled by an adaptive procedure[15]. For each persistence level, we ran two adaptive tests (30 trials each) targeting the 85% correct threshold speed, with an upper limit of 300 words per minute (wpm; corresponding to display duration of 0.2 s per word, which barely allows one saccade[16]) and a lower limit of 75 wpm (0.8 s per word). After each trial, participants would also report whether they saw strobing through a yes/no prompt. Three practice trials were performed with the lowest reading speed at the beginning of the experiment to familiarize participants with the task.

2.4 Eye movement analysis

The eye movements data were processed offline using customized functions in MATLAB (R2021a, MathWorks, Natick, MA). Eye position data were filtered with a second-order Butterworth filter (see Fig. 1B for an example), with cutoff frequencies (30-75Hz) individually decided based on per participant signal quality and visual inspection. Eye velocity was filtered with a 5-point Gaussian filter (corresponding to a time window of 20.8 ms). Saccades are detected based on combined velocity (70 deg/s) and acceleration (3000 deg/s^2) criteria and onsets/offsets computed via linear regression of neighboring points to find the time intercept, as done in previous work[17]. We focused the analysis on "word-switching" saccades, defined as saccades that occurred before fixating the last word, were at least 8 deg in amplitude, and crossed the midline of the wordlist display. To further exclude saccade outliers, we fitted a power function (peak velocity = K*amplitude^L)[18] for saccade main sequence[19] (which describes the consistent relationship between saccade amplitude, peak velocity, and duration) and excluded saccades that are outside of the 95% confidence interval (2.63 ± 2.79 across participants). Median saccade amplitude and peak velocity, and the fitted parameters K and L of the main sequence were compared across persistence conditions.

3. Results

The main goal of the study was to examine the effect of persistence on reading task performance, measured as the 85% threshold of the silent reading speed in the saccadic RSVP task. Since lower persistence was hypothesized to result in a stronger perception of strobing and larger saccadic undershooting errors[2], we predicted that reading speed would decrease with persistence. To test the effect of persistence on reading speed, as well as on perception of strobing and saccadic metrics, for each dependent variable we conducted a repeated measures analysis of variance (ANOVA) with persistence as the main factor.

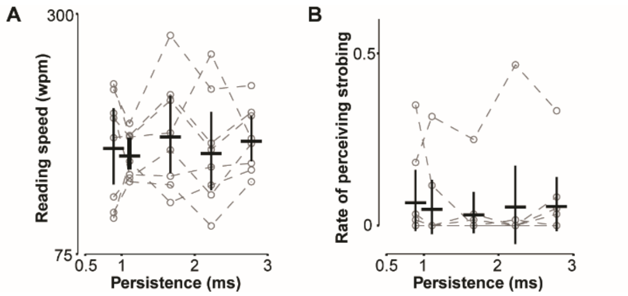

Figure 2. *(A) Reading speed across different persistence levels. The horizontal bar indicated the group average, and the error bar indicated the 95% confidence interval (CI). Grey dashed lines with open circles showed individual data. (B) Rate of perceiving strobing across persistence levels, shown in the same format as (A).*

3.1 Reading speed and perception of strobing

We did not observe a significant effect of persistence on reading speed (Fig. 2A; $F(4,28)=0.72$, $p=0.58$). There was no consistent trend across participants. The effect of persistence on strobing was also not significant (Fig. 2B; $F(4,28)=0.30$, $p=0.87$), likely due to the individual variability. There is known large inter-participant variability in the perception of strobing, with some people not seeing it at all, and others perceiving it more regularly at lower persistence levels[2], which is consistent with the pattern observed with most participants in the current study (Fig. 2B). However, one participant reported seeing strobing at a relatively high frequency which seemed to even increase with persistence levels.

3.2 Oculomotor metrics

To examine the effect of persistence on saccadic targeting, we analyzed word-switching saccades, which were expected to have similar magnitudes across trials. However, as lower persistence is correlated with larger undershooting error of saccades[2], we expected to see a decrease in saccade amplitude with persistence levels. Due to the consistent relationship of saccade main sequence, we expected peak velocity to be affected similarly by persistence. The fitted parameters K and L of the main sequence were also examined, as they represent the scaling of the relationship between saccade amplitude and peak velocity, and might be affected as well.

Surprisingly[2], we did not observe any significant effects of persistence on either saccade amplitude (Fig. 3A; $F(4,28)=0.24$, $p=0.91$) or peak velocity (Fig. 3B; $F(4,28)=1.03$, $p=0.41$). No significant effects were observed with fitted K ($F(4,28)=0.52$, $p=0.72$) or L ($F(4,28)=0.24$, $p=0.92$) either. To ensure that time pressure did not influence the effect, we redid the analyses with trials in which the word display duration was ±100ms around threshold individually. Still, we found no significant effects of persistence on saccade amplitude ($F(4,28)=0.91$, $p=0.47$) or peak velocity ($F(4,28)=1.91$, $p=0.14$). We examined one potential explanation for the lack of an oculomotor effect in the next section.

3.3 Spatial frequency of the stimuli display

To explore whether image statistics such as SF might have modulated the effect of persistence on saccadic control, we compared the image statistics of our stimuli to the stimuli in the previous study showing an effect of persistence on saccadic control[2] (Fig. 4A). Specifically, we conducted a Fourier analysis on the 2D images, then computed the 1D radial average of the power spectrum across SF channels. Two major differences could be observed (Fig. 4B): First, stimuli in the current study lack power at SFs >7 cycles per deg (cpd), partly due to the lower resolution and limitations of the optical modulation transfer function (MTF) in

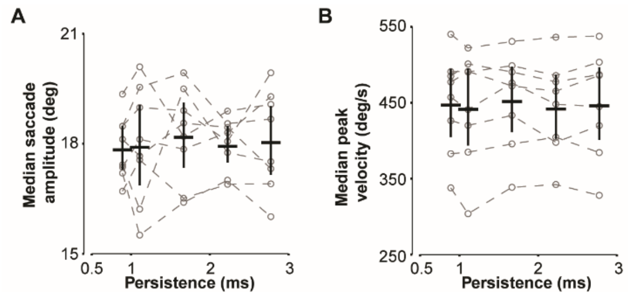

Figure 3. *(A) Median saccade amplitude and (B) peak velocity across persistence levels, shown in the same format as Fig.2A.*

Rift CV1, whereas the stimuli in the previous study had a nyquist SF limit of 12 cpd (with improved MTF). Second, the overall power in the current study, especially at SFs of 2-4 cpd near the typical peak of contrast sensitivity[20], is also smaller than the previous study, likely due to the fact that we have a uniform background. We discuss this limitation in more detail in the Discussion section.

4. Discussion

In the current study, we aimed to quantify effects of persistence on reading performance and potentially correlate them with perceived strobing and/or oculomotor deficits, with the ultimate goal of informing HMD persistence settings. In our novel saccadic RSVP task, we did not observe any effects of persistence on reading. Further examination of the perceptual artifacts and oculomotor metrics implied several potential limitations of current off-the-shelf display hardware, which could mask persistence driven effects. We discuss these limitations below, then discuss implications for future display technologies for VR and AR.

First, the off-the-shelf HMD used in the current study is prone to other spatial artifacts unrelated to persistence, which could mask temporally-driven effects via image degradation. For example, perceptual artifacts such as "god rays", off axis blur and/or chromatic aberration may arise in Rift CV1 due to its lens structure. Such artifacts could potentially impair image quality – especially with our high contrast stimuli – and may therefore limit the sensitivity of our task. Additionally, our ability to detect small changes in saccadic targeting[2] was limited by the spatial and temporal precision of our eye tracker – which was approximately 2x worse in both sampling frequency and calibration error spread than in the previous study. Newer and future generations of HMDs promise to have improved optics that limit off-axis blur and chromatic aberrations; it remains to be seen how low persistence artifacts and oculomotor deficits would manifest in these HMDs.

Second, the lack of an observed oculomotor deficit under low persistence levels could be due to the stimulus image statistics. One possible cause of eye movement errors in previous work[2] is the interruption of saccadic planning due to increased uncertainty of

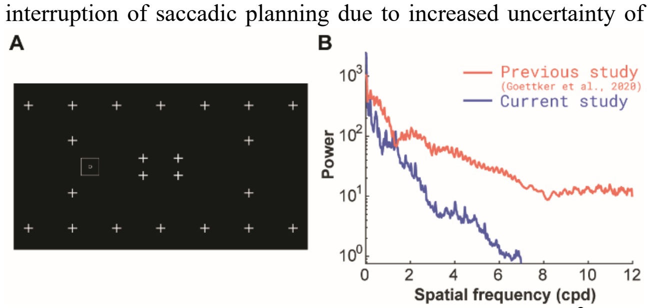

Figure 4. *(A) Stimuli display from the previous study[2] used for the Fourier analysis. (B) Theoretical SF power spectrum for stimuli in the previous and current study apparatuses.*

neural target representation under lower persistence levels. We postulate that image statistics (e.g., contrast energy across SF) could also contribute to this representation, thereby increasing the signal-to-noise ratio of the neural representation and eliminating targeting deficits. In the current study, due to a narrower range and lower spectral power of SFs especially around peak contrast sensitivity, the effect of persistence on both strobing visibility and saccadic errors might be attenuated. The lower overall luminance in the current study (restricted by highest luminance possible in the lowest persistence level after calibration; see Methods) could also contribute to a smaller persistence effect, with a reduction in the neural oculomotor signal strength[21], and reduced overall spatiotemporal contrast sensitivity[22]. Another possibility is that high SF information, which is lacking in the current study, contributes more to visible temporal light artifacts than low SFs[23]. There appears to be an interaction between the temporal and spatial display properties and the appearance of saccade-related artifacts.

As HMD technology advances to include wider ranges of SFs (through improved angular resolution and optical quality) presented across wider fields of view, we expect persistence-driven artifacts to play a more significant role in perceived quality and task performance – especially given current temporal HMD limitations. It is important to continue to address this tradespace as the technology develops and text-rich use case expectations arise in VR and AR (e.g., remote work/productivity). Future studies with improved hardware and flexible manipulation of image statistics are therefore needed to better inform a persistence optimization range based on task performance.

5. Impacts

- We designed a novel saccadic RSVP task to quantify the effect of HMD temporal properties on reading performance.
- Surprisingly[2], we found no effects on eye movement metrics and reading speed across persistence levels.
- Limitations in the hardware and stimulus properties in the current study provide directions for future studies to improve persistence optimization based on performance metrics for future VR/AR HMDs with improved display properties (e.g. spatial resolution, brightness, field of view).

6. Acknowledgments

Thanks to Saturnia Bocast and Helen Ayele for the help with data collection. As well, we thank Ian Erkelens, Kevin Rio, and Minjung Kim for help with the experimental design. Finally, we thank the excellent engineering support at Reality Labs at Meta for their help in preparing the experimental apparatus.

7. References

1. Murdison TS, McIntosh C, Hillis J, MacKenzie KJ. 3-1: Psychophysical Evaluation of Persistence- and Frequency-Limited Displays for Virtual and Augmented Reality. SID Symposium Digest of Technical Papers. 2019 Jun;50(1):1–4.
2. Goettker A, MacKenzie KJ, Murdison TS. Differences between oculomotor and perceptual artifacts for temporally limited head mounted displays. J Soc Inf Display. 2020 Jun;28(6):509–19.
3. Hershberger WA, Jordan JS. The Phantom Array: A Perisaccadic Illusion of Visual Direction. Psychol Rec. 1998 Jan 1;48(1):21–32.
4. Park S, Lee C-S, Kang H, Pak H, Wilkins A. Visibility of the phantom array effect according to luminance, chromaticity and geometry. Lighting Research & Technology. 2019;52(3):377–88.
5. Watson AB. A formula for human retinal ganglion cell receptive field density as a function of visual field location. Journal of Vision. 2014 Jun 30;14(7):1–17.
6. Jorna GC, Snyder HL. Image Quality Determines Differences in Reading Performance and Perceived Image Quality with CRT and Hard-Copy Displays. Hum Factors. 1991 Aug 1;33(4):459–69.
7. Vitu F, O'Regan JK, Mittau M. Optimal landing position in reading isolated words and continuous text. Perception & Psychophysics. 1990 Nov;47(6):583–600.
8. Clark JJ, O'Regan JK. Word ambiguity and the optimal viewing position in reading. Vision Research. 1999 Feb;39(4):843–57.
9. Pelli DG, Tillman KA, Freeman J, Su M, Berger TD, Majaj NJ. Crowding and eccentricity determine reading rate. Journal of Vision. 2007 Oct 26;7(2):1–36.
10. Näsänen R, Ojanpää H. Effect of image contrast and sharpness on visual search for computer icons. Displays. 2003 Oct 1;24(3):137–44.
11. Näsänen R, Ojanpää H, Kojo I. Effect of stimulus contrast on performance and eye movements in visual search. Vision Research. 2001 Jun 1;41(14):1817–24.
12. Blanc-Goldhammer DR, MacKenzie KJ. The effects of natural scene statistics on text readability in additive displays. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2018 Sep 1;62(1):1281–5.
13. Grainger J, Tydgat I, Isselé J. Crowding affects letters and symbols differently. Journal of Experimental Psychology: Human Perception and Performance. 2010;36(3):673–88.
14. Smith T, Guild J. The C.I.E. colorimetric standards and their use. Trans Opt Soc. 1931 Jan;33(3):73–134.
15. Owen L, Browder J, Letham B, Stocek G, Tymms C, Shvartsman M. Adaptive Nonparametric Psychophysics. arXiv:210409549 [q-bio, stat] [Internet]. 2021 Apr 19 [cited 2021 Sep 1].
16. Rayner K, Slowiaczek ML, Clifton C, Bertera JH. Latency of sequential eye movements: Implications for reading. Journal of Experimental Psychology: Human Perception and Performance. 1983;9(6):912–22.
17. Blohm G, Lefèvre P. Visuomotor Velocity Transformations for Smooth Pursuit Eye Movements. Journal of Neurophysiology. 2010 Oct 1;104(4):2103–15.
18. Dowiasch S, Marx S, Einhäuser W, Bremmer F. Effects of aging on eye movements in the real world. Frontiers in Human Neuroscience. 2015;9(46):1-12.
19. Bahill AT, Clark MR, Stark L. The main sequence, a tool for studying human eye movements. Mathematical Biosciences. 1975 Jan 1;24(3):191–204.
20. Barten PGJ. Formula for the contrast sensitivity of the human eye. Image Quality and System Performance. 2003 Dec 18;5294:231–8.
21. Bell AH, Meredith MA, Van Opstal AJ, Munoz DP. Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. Exp Brain Res. 2006 Sep 1;174(1):53–9.
22. Watson AB, Ahumada AJ. The pyramid of visibility. Electronic Imaging. 2016 Feb 14;2016(16):1–6.
23. Davis J, Hsieh Y-H, Lee H-C. Humans perceive flicker artifacts at 500 Hz. Sci Rep. 2015 Feb 3;5(7861):1-4.