Vergence-Accommodation Conflicts in Augmented Reality: Impacts on Perceived Image Quality

Ian M Erkelens*, Kevin J MacKenzie*

***Facebook, Redmond, WA**

Abstract

This study quantifies the 'Zone of Clear Vision' (ZoCV), which defines the magnitude of a vergence-accommodation conflict (VAC) that a user can accept in a binocular augmented reality environment before there is a perceived impact on image quality. Results indicate that the ZoCV extends up to 0.5 diopters on either side of a fixed focus display. This data correlates well to the Zone of Comfort, established from VR systems and suggests that an impact of perceived image quality may predict the buildup of visual discomfort overtime. Further, a subset of participants reported an impact of image quality on real-world content when simultaneously viewed with virtual content rendered with VAC, suggesting that rendered AR content outside the ZoCV can inadvertently impact some users view of the world.

Author Keywords

Vergence, Accommodation, Cue Conflict, Image Quality, Augmented Reality, Virtual Reality, Head Mounted Displays, Near Eye Displays

1. Introduction

Fixed focus stereoscopic displays induce a conflict in the stimuli to ocular vergence (gaze angle alignment) and accommodation (focusing) by rendering variable disparity with a fixed focus cue (Figure 1). The impact of this display artifact on perceived image quality in the context of additive display environments such as augmented reality (AR) systems is currently unknown.

The vergence and accommodation oculomotor systems are synkinetically linked (1) - where a stimulus to one drives a change in both. Therefore, changes in ocular vergence can drive an accommodative response away from a stereoscopic display's fixed focal plane - resulting in optical defocus, and potentially perceived blur. As a result, the user is required to decouple these oculomotor responses in order to maintain clear, single binocular vision. This viewing situation has implications on visual discomfort and task performance when identifying stereoscopically presented stimuli (2–4). In virtual reality (VR) systems, where the vergence-accommodation cue conflict (VAC) is global across the scene (Figure 1B), there is a range of content distances surrounding a given fixed focal display distance (FFDD) in depth where content can be viewed comfortably, known as the 'Zone of Comfort' (3).

In contrast, binocular augmented reality systems are additive with the real world and thus only rendered content will have VAC and thus vergence and accommodation must remain coupled when viewing real world objects and then decouple when viewing digitally rendered content on a fixed image plane. In many situations the real world and rendered stimuli can be at the same spatial location in depth, while having different focal cues (Figure 1 C&D). This places extra demands on the user relative to VR systems, where vergence must remain fixed and accommodation must vary between the two focal cues in AR. It is unknown whether the visual system can achieve these two different oculomotor control states (coupled and uncoupled) and, if so, how rapidly it can alternate between each. At a more basic level, the magnitude at which the difference between real world and rendered content focal cues begins to impact image quality of the rendered or real-world content is unknown. The goal of this study was to directly quantify the blur detection thresholds for high spatial frequency text rendered simultaneously at the same spatial distance with and without VAC of varying magnitudes in order to establish the 'Zone of Clear Vision'

2. Methods

Participants: 11 healthy participants with clinically normal binocular vision completed the study. This was defined as monocular visual acuity greater than 20/25, global stereopsis of at least 60 arcseconds. Participants heterophoria was also measured at 6 meters and 40 centimeters using the Modified Thorington Technique.

Apparatus: Stimuli were presented in an additive multiplane (3) binocular haploscope (5). This apparatus allows for up to 3 sets of content to be rendered at the same vergence distance, each with its own unique focal distance. We tested 3 FFDD for rendering the 'virtual' AR content; 0.5D, 1D and 1.67D.

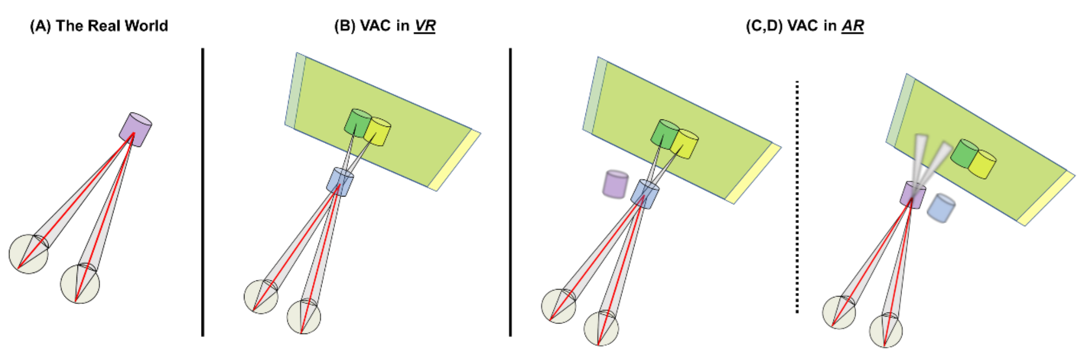

Figure 1: Illustration of the real world (A) and vergence-accommodation conflicts in VR (B) and AR (C/D). Ocular alignment (vergence) is defined by the red lines, ocular focusing (accommodation) by the grey elongated triangles. The two overlapped displays represent the dichoptic image presentation to the left and right eyes independently. Rendered objects (with VAC) are depicted by the blue cylinder and real-world objects as the purple cylinder.

Stimuli & Task: The task consisted of a brief presentation (4 seconds) of 2 lines of identical text rendered above and below the central fixation point. Both lines were rendered at the same spatial distance (identical disparity cue as in Figure 1 C&D), which was sampled linearly in dioptric space around the FFDD. One line of text had no VAC ('real-world' content) while the other had a fixed focal cue based on the 3 different FFDD tested ('virtual' content, with VAC). If the text was rendered in front of the FFDD, as in Figure 1 C&D, we refer to this as 'positive VAC'. Likewise, when the VAC text was rendered at a depth located behind the FFD distance for that condition, we refer to this as a 'negative VAC'.

Each set of text contained 8 Sloan font (6) single capital letters with equal spacing (0.5 letter width). Each line of text was surrounded with flankers to maximize crowding effects and was 20 arcmins in height, roughly equivalent to 20/60 Snellen font size and a spatial frequency of 7.5cpd. The text was white rendered on a black background (25:1 contrast ratio, maximum luminance 30cd/m^2).

Participants were positioned in the apparatus using a bite-bar and instructed to select which line appeared most clear to them with a button press (one interval 2-alternate-forced-choice task). There was a total of 42 different conditions and 12 trials were completed for each (504 trials total). The order of all 504 possible trials was randomized for each participant.

Data Analysis: If the participant selected the non-VAC content, that trial was given a score of 1 (0 if VAC content selected). For each condition the mean of the 12 trials was calculated and defined their preference score for a given text type. Based on the number of trials per condition, a binomial test of significance indicates that a preference score below 0.25 or above 0.75 is statically significant at $p<0.05$. This statistical criterion was used to define the blur detection threshold for VAC magnitude (Figure 2). The values were then normalized using equation (1) to derive a normalized probability of blur detection score with a detection threshold of 0.5 on this normalized scale. The Zone of Clear Vision (ZoCV) was then defined as the width between interpolated points where the mean blur detection is subthreshold on either side of the virtual content fixed focal distance.

$$\textit{Normalized Score} = |(\textit{Preference Score} - 0.5)| * 2 \qquad (1)$$

3. Results

Figure 2 and 3 illustrate the first main result of this study; the population level width of the Zone of Clear Vision spans 0.52D (±0.21) in front and 0.47D (±0.13) behind a given FFDD. This is consistent with previous work in terms of visual discomfort and VAC established by Banks et al. for fixed focus stereoscopic viewing (3). What is notable from Figure 3, is that the width of the ZoCV is narrower for the most distant focal distance tested (0.5D or 2m), especially for content rendered behind the FFDD. There is also greater variability in the width of the ZoCV for VAC content rendered in front of the display focal plane.

The second significant result can be observed in the top row of Figure 2, where the 3 participants plotted exhibited very different content preferences. One subject (Figure 2 top row, solid yellow lines) selected non-VAC text as clearer, regardless of VAC direction, which was our general expectation for this study. However, the other two participants plotted here demonstrated asymmetric text preferences, depending on the direction of the VAC. In Figure 2, one observer demonstrates a preference for VAC content being clearer when VAC content is positive (in front of the display focal distance) and the opposite preference (non-

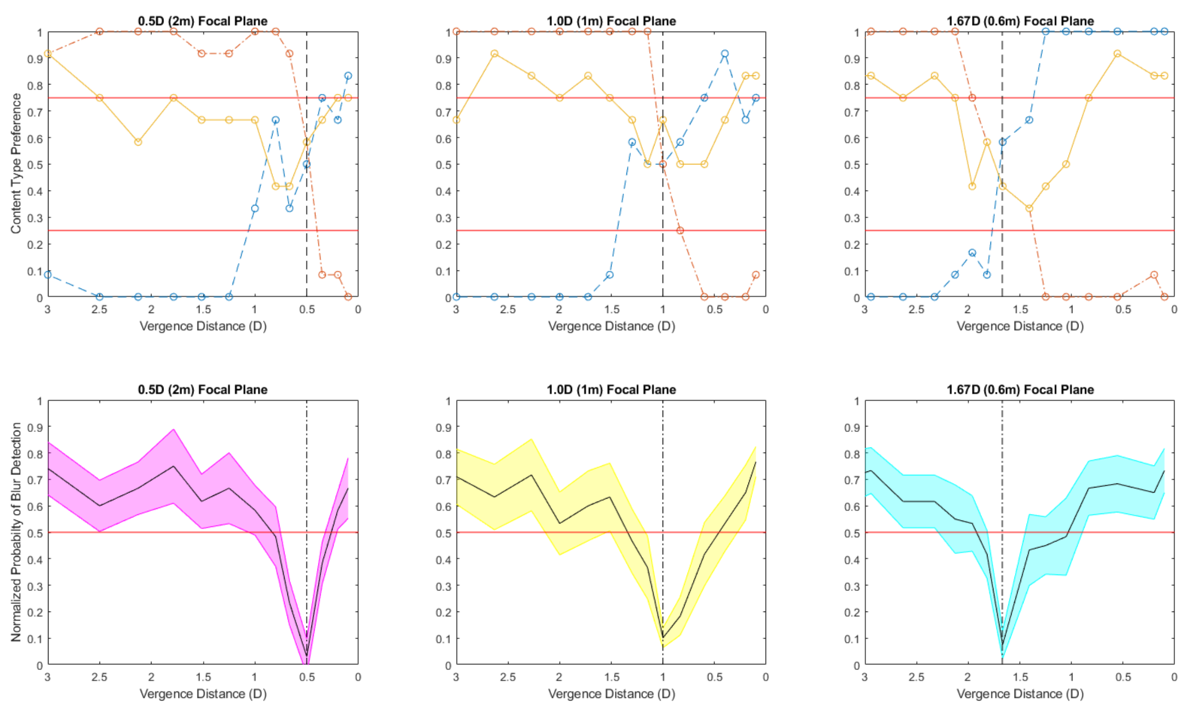

Figure 2: Raw preference scores for three different participants for each of the fixed focus conditions denoted by the vertical black dotted line (Top). Each data points x-value represents the vergence distance all the text was rendered at. Horizontal red broken lines indicate the point at which an individual's preference score was statistically significantly ($p < 0.05$). The bottom row of figures are the group-level mean (black line) and standard error (shaded areas) of the normalized blur detection scores. The horizontal red line indicates the detection threshold.

VAC content clearer) when the VAC direction is negative (rendered behind the displays focal distance - blue dotted line). The third observer plotted in Figure 2 demonstrates the inverse of this asymmetric preference behavior. It is also clear that the preference behavior was consistent for these 3 participants shown in Figure 2, regardless of the FFDD. Of the 11 subjects tested, 8 showed a consistent VAC directional preference for all FFDD tested, while 3 others had varying preferences depending on the FFDD.

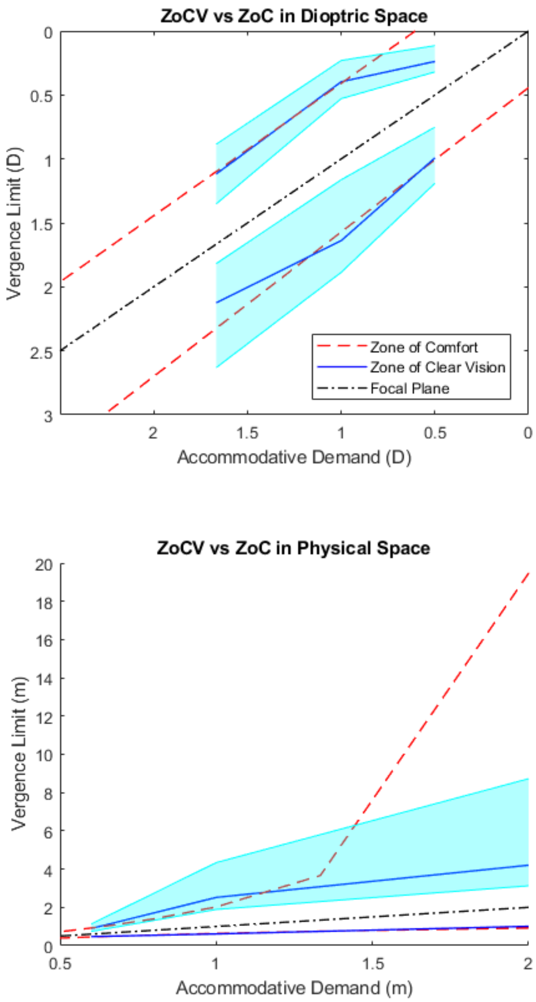

Figure 3: The Zone of Clear Vision (solid blue line – mean, shaded area – SE) plotted over the Zone of Comfort (ZoC data from Shibata; dotted red line). The broken black line represents equal vergence and accommodation demands (no VAC). The top figure plots the data in diopters while the bottom panel is the inverse, plotted in physical space (meters). Only the SE for the negative VAC condition is plotted in the bottom figure.

Overall, there were 3 participants that indicated the VAC text was always clearer when the direction of the VAC was negative (behind the FFDD) and had the opposite preference (non-VAC text clearer) when the VAC direction was positive. There were 2 participants with the inverse asymmetric preference to this (VAC text clearer for positive VAC conditions and non-VAC text clearer under negative VAC conditions). Only one participant reported the VAC content was always clearer, regardless of direction, and 2 that reported the non-VAC text was always clearer. The remaining 3 participants had a preference that varied based on the FFDD.

We investigated the role that heterophoria played in these unique preference observations. Heterophoria defines the vergence angle in the absence of retinal disparity information (monocular vergence angle at a given distance). To do this we classified the content preference type of each individual numerically (4 possible types) and then compared this with the expected heterophoria type (esophoria, exophoria or orthophoria) at each FFDD distance tested using a Wilcoxon Signed-Rank test. Because heterophoria was only measured at 6 meters and 0.33 meters, we had to interpolate the expected heterophoria for a given FFDD with a simple linear regression function. Orthophoria was defined as a heterophoria magnitude of ±1 prism diopters. Results of this additional analysis did not reveal a significant relationship between heterophoria type and content preference type for any of the 3 FFDD's tested ($p > 0.05$).

Figure 3 demonstrates the final significant result of this study; the width of the ZoCV is much larger in physical space when the FFDD is more distal to the user. Furthermore, as the FFDD is progressively moved more proximal to the user, there are only small gains in physical distances in front of the FFDD where content can be rendered without impacting perceived image quality while sacrificing a much larger amount of physical space for rendering content behind the FFDD.

4. Discussion

The data provide the first psychophysical assessment of the impact of vergence accommodation conflicts on perceived image quality in the context of additive display environments, specifically augmented reality binocular head mounted displays. The results follow similar observations regarding the impact of VAC on visual comfort and performance in VR systems and highlight the asymmetric width of these zones in physical space.

The most striking result was unexpected; a significant proportion (72%) of the population preferred to focus on the VAC text for one direction of VAC for at least one FFDD. This observation could be related, at least in part, to the sparseness of the real-world (non-VAC) content that was rendered in this study. One would generally expect a denser real-world background environment with very different scene statistics in a typical additive AR experience than what was used in this study. Such an environment may then drive the ocular focusing mechanism more towards the real-world non-VAC content than in this study, as new research suggests that peripheral defocus cues also drive accommodative responses (7). Thus, while our study design may not generalize to all potential AR use cases, the results demonstrate that VAC-related artifacts may actually impact the user's perception of the real world, as well as the rendered content under the conditions tested.

The preference for non-VAC content was not well explained by the users heterophoria. Another factor that could have influenced the preference behavior is the resting focus of the accommodative system, known as the 'dark focus'. The dark focus is an analogous accommodative measure to the resting vergence posture (heterophoria) and may act as an anchor to the focusing response when the user is presented with conflicting focusing cues between spatially congruent objects. This would explain some of the user's preference for negative VAC content for the most distant FFDD (0.5D), as the dark focus typically ranges between 0.25D and 3.0D (8). In this instance, for content rendered spatially beyond the FFDD, the accommodative response may be drawn towards the VAC contents focal distance, which is more congruent with the dark focus. The variance in dark foci between individuals may also explain the variance in the content preference behavior in this study; however, more work is needed to understand the influence of both heterophoria, dark focus and the dynamic interactions (strength of the cross-coupling) between vergence and accommodation to help pinpoint the exact driver of preference in a given individual.

It is unclear from the current data whether the text selected was actually 'clear' to the participants as the task required the observer to report which image was more clear, thus the outcome metric is relative to some baseline image quality, which may or may not be impacted by the amplitude and direction of VAC present. Previous work suggests that accommodation will strive to equalize contrast energy when presented with VAC conditions in the context of VR space (9). Without a direct measure of pupil size (which would define the eyes natural depth of field) or the accommodative response of each observer during each condition it is not possible to distill the impact of each of these factors to the variability in the interobserver width of the ZoCV further.

Future work will seek to obtain accurate real-time measures of these oculomotor parameters as well as more detailed assessments of their vergence and accommodative states in in order to help provide better individualized predictions the impact of VAC in AR systems will have on a user and how this relates to the expected depth of field, which is likely the major limiting factor of the ZoCV as defined in this study.

5. Impact

The results of this study add 3 important pieces of information to the existing understanding of vergence-accommodation cue conflicts in fixed focal display architectures. Specifically, the results apply to augmented reality environments where rendered content is additive and VAC conditions are local only to the rendered content.

1) The Zone of Clear Vision overlaps almost completely with the Zone of Comfort in VR, suggesting that when the user perceives an impact on image quality due to VAC magnitude, one may expect this to extend to long term comfort impacts. Given the difficulty in obtaining valid measures of visual comfort directly from users, having reliable and less variable quantitative metrics that correlate with qualitative visual discomfort are useful when larger scale user studies are impractical or not possible.
2) The width of the ZoCV is roughly symmetric in dioptric space, with the exception of content rendered behind more distant fixed focal displays. When the width of the ZoCV is converted into physical space, more distal fixed focal displays have a greater area with which to rendered content without impacting image quality. Further, there is an asymmetric trade space as the display focus is moved more proximal to the user; sacrificing large areas of space behind the display where content can be rendered while gaining a much smaller clear rendering area in front of the display.
3) 72% of the datasets studied preferred the rendered, VAC content in at least one VAC direction. This finding is important to understanding the impact of VAC in AR. Specifically, there may be a subset of users where rendering augmented content with VAC draws their focus away from their current accommodative posture, thus resulting in the real world becoming blurred. This can have significant implications for user safety and overall experience in AR devices.

6. Acknowledgements

The authors would like to acknowledge the engineering and software development support provided by Yusufu Sulai, Olivier Mercier and Cliff Owen on the hardware platform used in this work. We are also indebted to Mark Opaliski for supporting data collection.

7. References

1. Schor CM. A dynamic model of cross-coupling between accommodation and convergence: simulations of step and frequency responses. Optom Vis Sci. 1992 Apr;69(4):258–69.
2. Hoffman DM, Girshick AR, Akeley K, Banks MS. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. J Vis. 2008 Mar 28;8(3):33.
3. Shibata T, Kim J, Hoffman DM, Banks MS. The zone of comfort: Predicting visual discomfort with stereo displays. J Vis . 2011;11(8):11–11.
4. Kim J, Kane D, Banks MS. The rate of change of vergence-accommodation conflict affects visual discomfort. Vision Res. 2014;105:159–65.
5. Mercier O, Sulai Y, Mackenzie K, Zannoli M, Hillis J, Nowrouzezahrai D, et al. Fast gaze-contingent optimal decompositions for multifocal displays. ACM Trans Graph. 2017;36(6):1–15.
6. Recommended stardard procedures for the clinical measurement and specification of visual acuity. Report of working group 39. Committee on vision. Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences., Advances in ophthalmology. 1980. Vol. 41 p. 103–48.
7. Labhishetty V, Cholewiak SA, Banks MS. Contributions of foveal and non-foveal retina to the human eye's focusing response. J Vis. 2019;19(12):1–15.
8. McBrien NA, Millodot M. The relationship between tonic accommodation and refractive error. Investig Ophthalmol Vis Sci. 1987;28(6):997–1004.
9. MacKenzie KJ, Watt SJ. The stimulus to accommodation: Changes in retinal contrast matter, not the spatial frequency content. J Vis. 2011;11(11):516–516.