Gaze Direction in Virtual Reality Using Illumination Modulation and Sound

Eli Ben-Joseph and Eric Greenstein
Stanford EE 267, Virtual Reality, Course Report, Instructors: Gordon Wetzstein and Robert Konrad

Abstract

Gaze guidance is an important topic in the emerging field of virtual reality, where content creators have less control over where a user is looking, and maintaining the immersive experience is critical. An effective method of subtle gaze guidance would allow content-creators to better tell their stories without disrupting the user's experience. In this paper, a user study was conducted to explore how intensity modulation (flicker) and 3D sound can affect gaze direction. It was found that flicker was the most effective method of gaze direction and that sound had no significant effect on gaze direction. While these results were encouraging, more data is needed to determine whether the results are statistically significant.

1 Introduction

Unlike the traditional 2D environments in which a user is limited to viewing content on the screen in front of them, virtual reality (VR) brings the user into an immersive, 3D setting. Though virtual reality allows for a much more dynamic experience, content-creators now have less control over what the user is viewing in any particular scene. Storytelling, especially in gaming, is essential to the experience and plot. Unlike 2D displays, where content is pushed to the user, in VR the user can explore the scene as they please, making the job of a content-creator more difficult.

Given the plethora of distractions a user may face in a VR scene, how can a content-creator ensure that a user is looking at the right place in the scene to continue the story, but not ruin the immersive experience with obvious cues? Seamless gaze-direction is a problem that VR developers face: whether it be guiding a user to a virtual point of interest, or ensuring they do not hit a wall in reality, this is a problem that is yet to be solved. While gaze guidance has been studied on displays, there has been almost no research on its applications in virtual reality.

To approach this problem, we attempted to guide a user's gaze with two strategies: a very subtle illumination modulation (flicker) in the periphery of the field of vision, or a 3D-localized sound. Flicker was chosen because previous research indicates that flicker can be successfully used to subconsciously guide a user's gaze across a 2D monitor. Similarly, sound was chosen because it can guide attention and it is possible within a VR environment to incorporate localized 3D sounds.

2 Related Work

2.1 Head Angle as a Proxy for Gaze Direction

Because we are not using an eye tracker within the HMD, we had to find a method for approximating where a user was looking within the scene. Based on previous studies, it seems that head angle is a suitable proxy for eye position. A team at Microsoft Research demonstrated that eyes tend to lead the head, but head angle catches up very quickly (within approximately 1 second) [Slaney et al. 2014]. Similarly, a group researching HCI determined that gaze can be intuited via head pose as well [Weidenbacher et al. 2006]. Another study on group dynamics looked to understand who a subject was looking at in a meeting. Eye tracking was not available to the researchers, so they used head angle as a proxy for whom the subject was looking at. Post-meeting interviews with the subjects and showed that head angle was a good substitute for gaze direction, maintaining an accuracy of 88.7% [Stiefelhagen and Zhu 2002]. Given that head angle seems to be a fair proxy for eye position, we decided that it was an acceptable way of tracking where a user was looking within our scene.

2.2 Gaze Guidance Using Illumination Modulation

In 2009, Bailey et al. introduced a novel technique to direct a viewer's gaze about an image [Bailey et al. 2009]. They used subtle modulation to attract the attention of a viewer, noting that peripheral vision responds to stimuli faster than foveal vision. By modulating regions of the scene in the periphery of a viewer's vision, they caused the viewer's eyes to saccade to that region. Luminance modulation and warm-cool modulation were chosen, as the human visual system is very sensitive to these changes [Spillmann and Werner 2012]. A few groups have applied this method for medical training and visual searches [Sridharan et al. 2012], [McNamara et al. 2008]. While this technique was successful in directing users to targets in complex images, it used active eye tracking

to detect when a viewer's eye was moving towards the modulating object, then stopping the modulation. This limits potential applications as it requires an eye tracker setup to be present. Other groups have used the fact that foveal vision is drawn to regions of sharp focus or high detail, and sharpened and blurred sections of images accordingly. However, this alters the overall appearance of the image, unlike the modulation-based methods.

In EE 367, we investigated using flicker to subtly guide a user's gaze. The modulation was visible in peripheral vision, which attracts attention, but invisible in foveal vision, so as to be subtle. This technique worked well on simple scenes, and we have since investigated how it can be used for search tasks. We believe that this method can successfully be extended into VR to guide a user's gaze.

2.3 Gaze Guidance Using Sound

The visual and auditory systems in the brain have their separate cortexes for the majority of information processing, yet are also very connected. In fact, there are certain types of bimodal cells which are responsible for this integration of different sensory inputs. Studies show that these bimodal cells are likely responsible for the connection between auditory cues and visual focus [Lewald et al. 2001]. Sound has also been shown to draw ocular attention [Glotin 2014], [Quigley et al. 2008], [Perrott et al. 1990], [Spence and Driver 1997]. However, the types and position of the sound are not all equal. Studies have shown that different types of sound have different success rates at actually drawing human attention (e.g. voice) [Song et al. 2013]. This information should be taken into consideration when creating our sound stimulus for gaze guidance.

2.4 Applications to VR

Though there are a few papers published in regards to eye-tracking systems within VR, there are no papers that study various ways to affect a user's head pose within a VR environment by using either sound or other visual techniques. This indicates that our field of research for this project is quite novel and relevant to some of the issues that content developers are facing today. In this experiment, we investigate how illumination modulation and 3D-localized sound can direct a user's gaze in virtual reality.

3 Method

3.1 Hardware and Software

All experiments were carried out on the head-mounted display (HMD) that we built during class. The screen

was a 6" 1080p LCD manufactured by Topfoison, and the IMU was a InvenSense MPU-9255. The ViewMaster VR starter pack and an Arduino Metro Mini (for IMU processing) were used. The test scene was created using Unity and packages that were downloaded from its asset store. Figure 1 shows the scene used. The experimental setup and design was mostly done in C# scripts attached to various objects in the Unity scene.

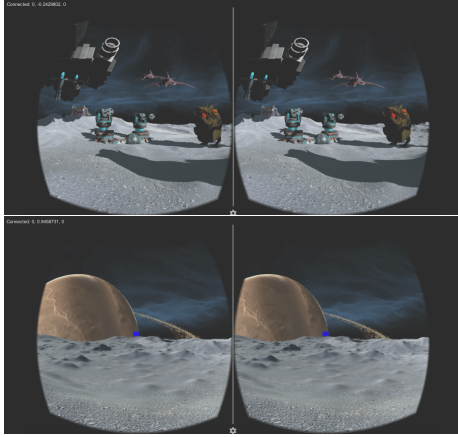**Figure 1:** *View of the experimental scene.*

3.2 Experimental Procedure

We tested 15 users overall. Users were selected to be in the control, flicker, or sound groups at random (5 per group overall). An object-of-interest (a flashing cube) was placed at a random position outside the field of view of the user, and the user was instructed to simply observe and explore the scene. Figure 2 shows the view of the scene from the user's perspective. The time it took for the user to notice the cube, as measured by the head orientation of the user, was recorded. As discussed in the relevant literature section, head angle is a good proxy for gaze direction, so when the cube was in the center of the field-of-view, the timer was stopped. After the experiment, we confirmed that the user did indeed see the object. Users were also asked about their age, vision history, perception of the quality of the scene, and what they thought about the sound and flicker.

3.3 Gaze Guidance Using Illumination Modulation

Our previous experiments conducted in EE 367 showed that if a certain segment of a display is modulated at the right frequency and in the right color-band (RGB), the modulation (flicker) is only visible in peripheral vision but not foveal vision. This is due to differences in the way eye cells are structured across the fovea and periphery. If the flicker is placed at the right position in the display, a user's gaze can be directed around simple scenes.

In our environment, flicker was implemented in the right

Figure 2: *Views of the experimental scene with the user facing forwards and backwards.*

Figure 3: *Screenshots of the flicker with the blue channel increased (top) and decreased (bottom).*
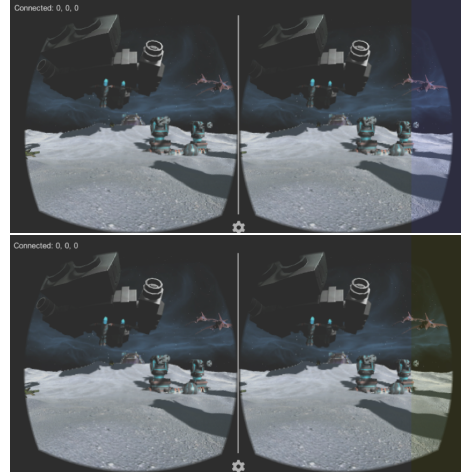
periphery of the user's right eye, with the goal of inducing a head turn in that direction. It was implemented by creating a UI canvas in Unity that sits over the entire image frame. This canvas is typically used for creating a HUD effect or intro menus, however we modified it to work for our flicker. To create the flicker effect, we covered the entire screen with a gray rectangle ($R = G = B = 130$) with an alpha value of 0.35 (so the scene could still be viewed behind it). Dimming the pixels uniformly should not result in a worse viewing experience, as in the isolated environment of virtual reality, the viewer's eyes will adjust under these new lighting conditions to perceive colors correctly.

On the right side of the screen, a rectangular band of pixels stretching the whole height of the screen had their blue channel value altered by $\pm 12\%$, so on average the modulating band was no different in color from the rest of the overlay. The flicker parameters were chosen based off our previous research so that the flicker would be slightly noticeable in the periphery and not in the fovea.

Within the Unity code, if we designated the flicker setting, we had the overlay switch between the low and high blue channel images, creating a unified image across the entire screen, but with the right side of the right eye region modulating. Figure 3 showcases this setup in more detail.

3.4 Gaze Guidance Using Sound

Previous research shows that human gaze is highly tied to auditory stimulus. When a more distinct sound is emanated, people tend to direct their gaze at the stim-ulus. This has ties to evolutionary advantage, as obtrusive sounds may have indicated a predator or enemy approaching (or a baby crying). With this in mind, we attempted to use an auditory stimulus to guide a user's gaze.

To do this, we attached a sound file to the cube in Unity, and used the built-in 3D sound model to compute stereo sound that adjusted for the user's position and orientation. A laser sound was chosen because it was fitting for our space battle scene. It played in a loop that repeated approximately every 2.5 seconds. A logarithmic volume roll-off with a maximum distance of 50 was used.

3.5 Data Analysis

Using the locations of the randomly-placed object and the user in the scene, we calculated the angle that the user would have to turn their head to view the object in their fovea. Using this information and the time it took the user to find the object, we calculated how many seconds it took the user to rotate their head to the object. Two-tailed t-tests were used to compare these time per angle measurements across the different groups. Time per angle was used as the metric for comparison because we wanted to adjust for how far the user had to rotate to see the object.

4 Results

The results of the experiment are shown in Figure 4 and Tables 1 and 2. Interestingly flicker was the most effec-
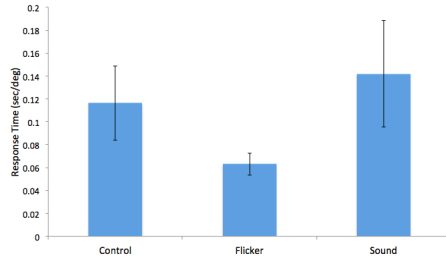
Figure 4: *Average response time to gaze-directing stimulus.*

	Average (sec/deg)	Standard Deviation
Control	0.1164	0.0725
Flicker	0.0632	0.0216
Sound	0.1418	0.1040

Table 1: *Comparison of different gaze directing techniques in terms of average response time.*

tive method of moving the viewer towards the object of interest, with an average of 0.0632 sec/deg. This was followed by the control group at 0.1164 sec/deg and the sound group at 0.1418 sec/deg.

The initial results are encouraging given the limited amount of data collected, but not yet conclusive. While the average time to find the cube was lowest for subjects in the flicker group, it only held a p-value of 0.18 when compared to the control group. With more data, we may find flicker to have no effect, or to have a statistically significant effect. The average time to find the cube was higher for subjects in the control and sound groups. Comparing the sound vs. control groups with a t-test yielded a p-value of 0.67, and comparing the sound vs. flicker groups yielded a p-value of 0.17.

After the experiments, we asked subjects about the quality of the images they saw, and how they perceived the sound and flicker. Two of the five subjects exposed to the flicker mentioned noticing the flicker in the right corner of their eye, which indicates that the flicker setup needs tuning to be made more subtle and effective. In previous experiments, the flicker was detected by a much smaller proportion of the population. One reason for the VR flicker being more noticeable is due to the low frame rate of the demo. We noticed that when the IMU was plugged in and the scene was running, the frame rate would drop from 50+ frames per second to 15 frames per second. As discovered in previous research, the flicker should ideally run at 30+ frames per second, requiring a monitor refresh rate of 60+ frames per second. Improving the hardware/software by putting the IMU on a dedicated

Comparison	P-value
Flicker vs. Control	0.18
Sound vs. Control	0.67
Flicker vs. Sound	0.17

Table 2: *Results of two-tailed t-test on average response times of the different groups.*

core/thread or using an Oculus Rift or other HMD that has optimized these sorts of problems would improve this issue.

It is also interesting to note that sound was polarizing to users; some users told us that they wanted to find the source of the sound, and fairly quickly located the object of interest, yet others mentioned that the sounds did not draw their interest at all. Another user struggled with the localization of the sound, having difficulty pinpointing what exact direction it came from. An improved spatial sound model or a more continuous sound that allowed for more rapid feedback perhaps could have improved the results. It is also likely that a fully surround sound system (rather than stereo headphones) would help with the localization. More importantly, the type of sound selected is very important, as literature suggests that humans react differently to various types of noises. In general, using sound is more intrusive than flicker, as users can always hear the sounds and they may thus disrupt the experience if not chosen wisely to fit the scene.

5 Future Work

There are numerous subjects to explore going forward. First and foremost, more data should be gathered across flicker, sound, and control to determine if the improvement brought on by the flicker holds a statistically significant advantage over the other methods, and if the sound strategy is indeed no different than control.

Within the flicker case specifically, more fine tuning of the parameters (size, color, and frequency) can be done to ensure that the flicker is subtle (so no subject notices it) yet still effective. Using a more state-of-the-art VR system would likely help, as their frame rates are higher than what we used, and their orientation tracking systems will be faster and more accurate.

The sound method also warrants further improvement. Literature shows that certain sounds work better than others at drawing attention; human voice in particular is a strong cue. It is also possible that certain frequencies and loop intervals are more effective than others.

In addition to flicker and sound, perhaps there are other cues that can be used to subtly direct a user around a

scene. For example, it is also possible that certain blurring effects can be used to guide a user's gaze. In conversation with industry experts, methods such as lighting (e.g. moving a light towards an object of interest) and motion of objects in the scene (e.g. a person walking across the user's field of view) are currently being used to improve game design, but could be explored in a more rigorous manner.

Acknowledgments

References

BAILEY, R., MCNAMARA, A., SUDARSANAM, N., AND GRIMM, C. 2009. Subtle gaze direction. *ACM Transactions on Graphics (TOG) 28*, 4, 100.

GLOTIN, M. H. 2014. *Effect of sound in videos on gaze: Contribution to audio-visual saliency modeling*. PhD thesis, Citeseer.

LEWALD, J., EHRENSTEIN, W. H., AND GUSKI, R. 2001. Spatio-temporal constraints for auditory–visual integration. *Behavioural brain research 121*, 1, 69–79.

MCNAMARA, A., BAILEY, R., AND GRIMM, C. 2008. Improving search task performance using subtle gaze direction. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, ACM, 51–56.

PERROTT, D. R., SABERI, K., BROWN, K., AND STRYBEL, T. Z. 1990. Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics 48*, 3, 214–226.

QUIGLEY, C., ONAT, S., HARDING, S., COOKE, M., AND KÖNIG, P. 2008. Audio-visual integration during overt visual attention. *Journal of Eye Movement Research 1*, 2, 1–17.

SLANEY, M., STOLCKE, A., AND HAKKANI-TÜR, D. 2014. The relation of eye gaze and face pose: Potential impact on speech recognition. In *Proceedings of the 16th International Conference on Multimodal Interaction*, ACM, 144–147.

SONG, G., PELLERIN, D., AND GRANJON, L. 2013. Different types of sounds influence gaze differently in videos. *Journal of Eye Movement Research 6*, 4, 1–13.

SPENCE, C., AND DRIVER, J. 1997. Audiovisual links in exogenous covert spatial orienting. *Perception & psychophysics 59*, 1, 1–22.

SPILLMANN, L., AND WERNER, J. S. 2012. *Visual perception: The neurophysiological foundations*. Elsevier.

SRIDHARAN, S., BAILEY, R., MCNAMARA, A., AND GRIMM, C. 2012. Subtle gaze manipulation for improved mammography training. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, 75–82.

STIEFELHAGEN, R., AND ZHU, J. 2002. Head orientation and gaze direction in meetings. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*, ACM, 858–859.

WEIDENBACHER, U., LAYHER, G., BAYERL, P., AND NEUMANN, H. 2006. Detection of head pose and gaze direction for human-computer interaction. In *Perception and interactive technologies*. Springer, 9–19.