Interaction Techniques Using Head Gaze for Virtual Reality

Rowel Atienza, Ryan Blonna, Maria Isabel Saludares, Joel Casimiro, and Vivencio Fuentes
Electrical and Electronics Engineering Institute, University of the Philippines
rowel@eee.upd.edu.ph

***Abstract*—Virtual Reality (VR) has finally caught the attention of major technology companies, giving it a significant boost towards becoming market-ready. VR has the capacity to create profoundly engaging experiences by synthesizing highly detailed sensory input in real time, making games the most obvious applications of VR. While the promises of VR may be exciting due to the problems it can solve and new experiences it can offer, there are still numerous issues to be addressed in order to achieve maximal immersion in VR. In this paper, we present the issues encountered during open testing of our applications on commercially available VR goggles and hand input device. The inability to effectively use the hands and feet to control and navigate the environment degrades the level of immersion of VR. With users preferring reliability over intuitiveness, we explore the use of gaze as a substitute input device. Results of experiments conducted on numerous volunteers showed significant improvement in terms of user control.**

Keywords*–*Virtual Reality; Head Gaze; Gesture Recognition; User Interaction.

Figure 1. Slash the Fruit game. The goal is to slash fruits as they are thrown at the player by monkeys. The slashing action is tracked by the Leap Motion device.

I. Introduction

Eye, head, and hand movements are well coordinated in everyday tasks [1]. It can therefore be said that head orientation as well as eye movements are precursors of the following manual actions – the head must first be oriented so that the eyes can locate an object's position, before the hands can grab it. This suggests that if head orientation can be reliably detected in certain contexts and tasks, it should be possible to approximate gaze and therefore predict the next manual actions even before they are performed [2]. In this paper, this postulation is used in the construction of a system of interaction with virtual environments. Head orientation as an approximation of gaze direction will hereafter be referred to as *head gaze*.

The first head mounted display (HMD) for virtual reality (VR) came into existence in the 60's [3] and – as with many other cutting-edge technologies – were limited to military applications. The device was bulky and sported low-resolution displays in its early form [4], which did little to encourage widespread adoption of the technology. Public interest in VR was revived when a head-mounted headset by Oculus became the product of a very popular Kickstarter project in 2012 [5]. Many major technology companies have since claimed their stake by investing in technologies related to Virtual Reality (VR), Augmented Reality (AR), or Mixed Reality (MR). As a matter of fact, NVIDIA – one of the leading manufacturers of graphics hardware – claims that the company has been maintaining close ties with at least 250 startups working on different aspects of VR/AR/MR [6].

VR has made progress by leaps and bounds in many of its problem domains, such as realism, responsiveness, device portability, and head tracking [7]. However, current VR control mechanisms are still in their infancy; they are either very bulky or require very specific hardware and software configurations [8]. The use of commercially available VR head-mounted displays (Oculus DK2) for visual feedback, as well as controller technologies such as the XBox controller and the Leap Motion controller [9], can help address this issue. To that effect, the authors have built VR applications targeting commercially available devices to allow for ease of replication.

The first attempt at building a VR game was to convert an existing 3D mobile game developed by some of the authors (available in the Apple App Store as *Slash The Fruit*, shown in Figure 1) into a VR game [10]. In contrast to employing a touchscreen for input, the Leap Motion controller was utilized to control a sword using three-dimensional hand movements. While users found being able to play in VR enjoyable, many were disappointed by the lack of reliability of the game controls. This was due to the fact that the Leap Motion has a very limited operational 3D space. Although replacing the Leap Motion with an XBox controller increased the reliability of input, it was far less intuitive and natural, thus severely degrading the VR experience.

In the next iterations, techniques to enable better in-game interaction with the Oculus VR were iteratively developed. Head gaze [11] was eventually discovered as a superior mode of input in the third iteration, which was further tweaked and refined in the last iteration.

Head gaze is now being used in the authors' demonstrations. This system is simple and accurate enough to be used in most common gyroscope-enabled devices such as smartphones. It was found in many tests that the simple head gaze was sufficiently intuitive to provide highly engaging and immersive VR gaming experiences.

In this paper, we present head gaze interaction applications and various insights for creating immersive VR games given the limitations of the current state of the art and commercially available devices from three game demonstrations: Slash the Fruit, The VREx, and Dungeon VR. The user experience

(a) Slash the Fruit setup. (b) Head gaze ().

Figure 2. Game setups.

Figure 3. (*top*) Control panel showing the options and (*bottom*) butterfly glowing when gazed at.

was evaluated by conducting a survey on several aspects of the demonstrations such as its graphics, controls, technical capabilities of head-mounted display, awareness in virtual world, and game experience.

II. Demonstrations

A. *Slash the Fruit VR*

Real movements and actions are analog, which contrasts with the digital nature of actions in modern games. RGB-Depth and infrared sensors such as the Leap Motion [9] attempt to bridge the gap between analog and digital interaction by directly translating hand and arm movements into computer-rendered digital models. These types of sensors are limited by their field of view and tracking accuracy [12], thus quickly degrading in consistency when used in non-ideal environments. For instance, the Leap Motion only has a field of view (FOV) of 150 degrees wide and 120 degrees deep, which limits tracking of movements to just beyond normal peripheral vision [13].

Slash the Fruit VR is a game that was presented in an early demonstration, where the participants were asked to stand before a Leap Motion controller (mounted on a tripod) while wearing the Oculus DK2 headset, as shown in Figure 2a. The Oculus DK2 headset is a VR head-mounted display (HMD), which serves as the primary visual feedback device for the player. The Leap Motion controller – a set of monochromatic IR cameras – allows detection of hand movement; actions can be mirrored in the game as a way to interact with the virtual world.

The game world is seen from the perspective of a sword-wielding avatar in a valley where several monkey are throwing fruits at him. To initiate interaction with the virtual sword, the player must perform a physical grabbing motion to simulate the action of gripping a real sword handle. Players can then proceed to mimic sword-fighting motions to achieve the in-game objective: to slash as many fruits as possible within the allotted 1.5-minute time limit. Physical slashing movements are directly translated by the Leap Motion into in-game avatar movements, allowing players the illusion of slashing fruits.

There were over 700 participants throughout the course of the demonstration, most of which were high school and college students between 13 and 25 years old. 85 participants were randomly selected and interviews were conducted to evaluate the game's reception. The participants were asked to rate (from 0 being the lowest to 5 being the highest) and comment on four aspects of the game: dizziness (participants were asked if they experienced dizziness during gameplay), graphics, control (ease of use), and overall gameplay experience.

Survey results showed an average rating of 2.0 for dizziness, 3.8 for graphics, 3.0 for control, and 4.2 for gameplay. It was interesting to note that most of the participants did not suffer from dizziness. The rare cases of dizziness resulted from improper fitting of the Oculus HMD, deposits on the HMD lens, or quick movements of the head. Some also complained that the HMD lenses could not "auto-adjust" to focus on the scene. While the idea of using the Leap Motion was appealing to many of the participants, issues were encountered on input's reliability. Many of the participants reported difficulty during attempts to manipulate an in-game sword.

B. *The Virtual Reality Experience (The VREx)*

The next iteration – informally dubbed *The VREx* or *The Virtual Reality Experience* – employed a different technique for interaction and visualization. This game added new aspects such as photo-realism, living creatures with reactive behavior, and control via head gaze (Figure 2b). To remedy the issue on input reliability of the Leap Motion, head gaze was instead used to interact with in-game user interface elements – a more natural alternative to using a mouse or console controllers in modern games [14]. The concepts of actions and verbs using head gaze were introduced: actions being the movements performed by the player, and verbs being the resultant actions performed in the game [15]. Head gaze served as the sole means of interaction with the virtual environment, effectively replacing the Leap Motion.

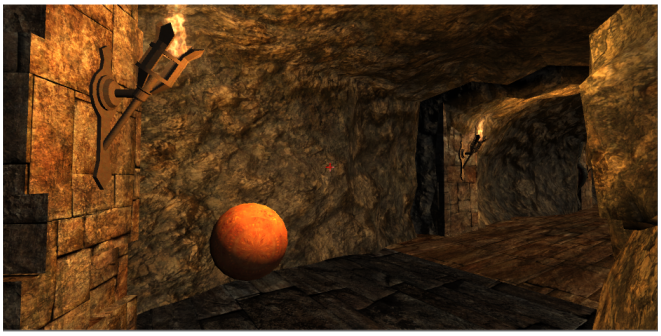

Figure 4. A marker in the Dungeon demonstration.

Accuracy and FOV in this case are limited only by the head-mounted VR device itself, because the players vision and interaction are directly correlated. Interaction is as simple as turning ones head in the direction of the interactive visual element, which then activates after a delay of 1 second and a visual/auditory cue. This familiar interface allows for a simple and intuitive control over many aspects of a game. Using appropriate delays, visual as well as auditory feedback, and non-obtrusive positioning of the interactive elements, the user's intention can be extrapolated and then executed in the proper fashion. The visual elements of this system can be easily molded to fit various in-game scenarios and uses as shown in Figures 3 and 4.

Floating text panels like those shown at the top of Figure 3 were used as interactive prompts to give the player a degree of control in the game. These panels were placed in designated "scenic" spots in the game map, and were comprised of two components: the immediately visible text prompt component, and a hidden "OK" button that appeared when the text prompt was gazed at. Gazing at the "OK" button when it appeared served as confirmation to proceed, causing the player to glide to the next area. Gazing away from the entire panel served as a cancellation gesture, causing the the "OK" button to disappear once again.

A fluttering butterfly (shown at the bottom of Figure 3) served as another of the visual elements for enabling traversal within the environment, used in the less modern-looking areas of the game. To compel players to observe the butterfly, it was programmed to fly around the player and perform interesting actions such as gracefully landing on nearby rocks and trees. To create visual feedback that encouraged the player to associate the act of gazing to an event, the butterfly would glow brightly the moment it crossed the player's line of sight. After a 1-second delay, the butterfly was set to emit a wind chime sound and cause the player to glide to a new location. It was observed that most players discovered the glowing effect of the butterfly almost immediately, and its capacity of transportation soon after.

The VREx demonstration showed that head gaze was an effective interaction device with the virtual environment. This was used to trigger a specific activity by looking at the selection panels, and to move to pre-set locations by looking at the butterfly.

(a) Avatar idle pose third-person perspective (left), and as seen by the player (right).

(b) Avatar walk animation third-person perspective (left), and as seen by the player (right).

(c) Avatar climb animation third-person perspective (left), and as seen by the player (right).

Figure 5. Dungeon avatar animations.

C. *Dungeon VR*

The latest application, named *Dungeon*, was designed to extend the core features built into *VREx*: head gaze as an interaction device, and photo-realism in graphics. Due to positive feedback on using head gaze in the previous tests, more verbs that could be activated using head gaze were added. Text prompts were removed for a more natural and integrated interface. Walking in-game is now possible by simply gazing at special floating markers. A climbing verb is automatically activated if obstacles are encountered during traversal between two markers. In contrast to *VREx*, the player's avatar is visible and animates correspondingly to the action being performed. This gives the player visual feedback, allows better visualization of the experience in the virtual environment, creates a stronger association between the player and the game, and reduces disorientation while performing various actions. This feature achieves an effect similar to that of modern first-person games such as *Crysis* [16].

Uniform in-game objects called "markers" serve as the interactive elements in the game. These objects are animated, which helps attract player attention and curiosity. Verbs performed when activating these markers are automatically determined depending on the context of the activation. A normal walking verb is performed when traveling on relatively flat ground, and a climbing verb is performed when an obstacle of sufficient size is encountered. This system is flexible in that it can be used with minimal adjustment in any type of virtual environment.

In *Dungeon*, the player avatar is visible and animated as shown in Figure 5. The avatar is programmed to perform animations based on the context of the current action, allowing for better visual feedback and in-game presence. Activation of

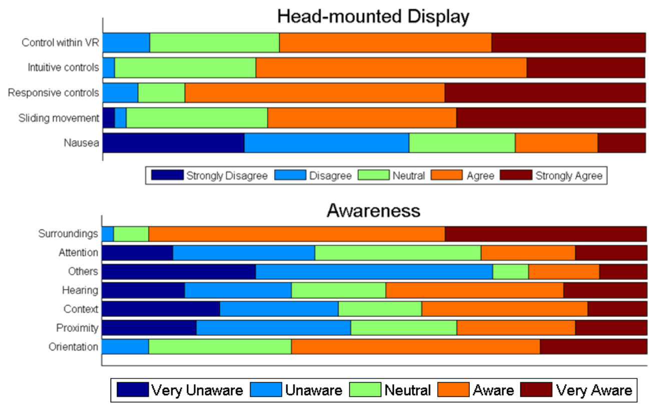

Figure 6. Survey response of users on: fidelity of virtual objects, technical capabilities of head-mounted display, and awareness in virtual world.

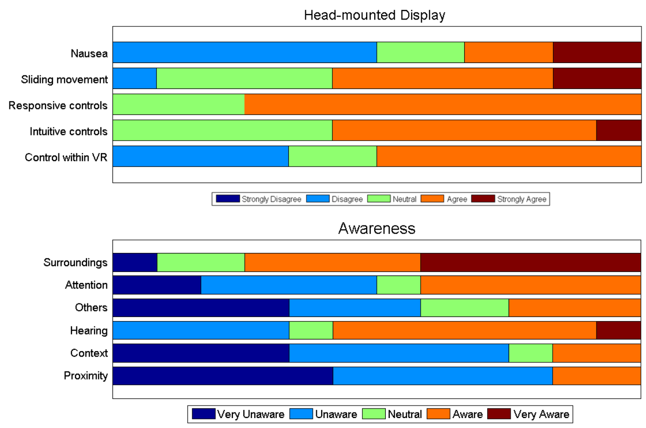

Figure 7. Survey response of users on: technical capabilities of head-mounted display, and awareness in virtual world.

the different animations in *Dungeon* is dependent on the speed and direction of player movement, as well as any collision events with the avatar. Horizontal motion causes the run/walk animation to activate, and smoothly interpolates based on avatar speed. A vertical motion component causes the avatar to activate its climbing animation. Vertical motion occurs each time a collision event is fired whose normal vector is facing the player.

The *Dungeon VR* demonstration expands the use of head gaze for interaction by introducing markers scattered throughout the dungeon to allow traversal of the virtual world, either by walking or climbing. This game also introduces a player avatar that moves within the virtual world. The presence of an avatar also gives a better experience of what it must feel like to actually be within the virtual environment, as opposed to being the disembodied observer in *VREx*.

The summary of the gaze actions with their corresponding verbs (effect in VR) used in the game demonstrations is shown in Table I.

	Action (Gaze at)	**Verb**
The VREx	Prompt panels	Glow, display text prompt "Ok"
	Butterfly	Glow, wind-chime sound, travel to next location
Dungeon VR	Marker on ground	Walk towards marker
	Marker on higher ground	Climb to marker

TABLE I. Gaze actions with their corresponding verbs of the game demonstrations.

III. Experiment

The user experience was evaluated by conducting a survey on several aspects of the demo: technical capabilities of head-mounted display, awareness in virtual world, and game experience [17]. Results of the surveys on ***The VREx*** and ***Dungeon*** are summarized in Figures 6, 7 and 8.

For ***The VREx***, 46 participants were asked to evaluate the 2-3 minute demonstration. There were three stages in the demo: futuristic present (set in a laboratory), past (Prehistoric landscape), and post-apocalypse (destroyed landscape). The technical capabilities of the head-mounted display focused on the ease and intuitiveness of controls in the virtual environment. It was observed that the head gaze was effective as a means of control in the environment. Without much assistance, the players were able to navigate the virtual world using this control. Several participants reported nausea due to the sliding as the movement mechanism. But majority deemed sliding as an agreeable means of movement. The awareness of participants in the virtual world was also evaluated. This area focused on identifying whether the experience was immersive enough to let the users feel that they are in the virtual world. Many of them felt that they were in the virtual world, however they were still able to hear what was happening outside of VR. It was also noted that they could not discern when someone is close to them or what exactly is happening in the real world once they were immersed in VR.

The ***Dungeon*** demonstration had 12 participants where the game lasted 8-12 minutes per play. There were two stages in the demo: exploration of some regions of the dungeon, and the maze leading to one of the missing gods. Another aspect that was tested was the intuitiveness of the controls implemented without game tutorials. The players were able to explore and navigate the virtual world using the head gaze directed to the markers without prior instructions. The earlier iteration focused on using the head gaze to trigger the movement of the player (looking at the butterfly) to the preset destination. In this iteration, the head gaze skill was expanded to execute different verbs (walking and climbing). The first stage allowed the user to explore the dungeon by gazing at the markers along the pathways. Depending on the structure of the dungeon, the user either walks or climbs to reach his next destination. They were able to complete the task of navigating the maze to find the missing god by using these verbs.

The game was evaluated by identifying whether the elements presented (such as virtual world, visual cues) created a compelling game to play with as shown in Figure 8. Most of the participants were neutral, agree, or strongly agree on the immersive environment and the helpfulness of the visual cues for exploration. All the participants completing the game without assistance demonstrated the intuitiveness of the head gaze as game control. More than half of the participants found it a compelling game while the rest were neutral or disagreed. Only less than half of the participants reported the game to be

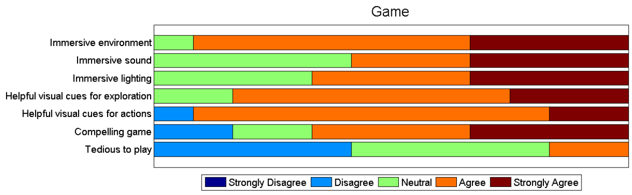

Figure 8. Survey response of users on game experience.

not tedious.

IV. Conclusion

We discussed the different issues and challenges encountered in virtual reality usability in several game demonstrations. ***Slash the Fruit VR*** introduced the use of Leap Motion to motion-capture slashing movements of the player and animate the avatar. However, due to the sensor's low input reliability, the following game/demo iterations used head gaze as the primary control method. ***The VREx*** showed that head gaze was an effective interaction device with the virtual environment to trigger a specific activity by looking at the selection panels, and to move to preset locations by looking at the butterfly. ***Dungeon*** game expanded the use of head gaze for interaction by introducing markers, and enabling player to walk or climb in traversing the virtual world. Consistency in the immersion provided by the VR experience was attained by introducing animals with various behaviors, enabling the avatar and its associated animations, and using head gaze to interact with the environment. Gestural gaze was mapped to elicit appropriate responses from the environment and its objects, and context-aware avatar animations were utilized. Experimental results and documented experiences of various users show the effectiveness of our proposed techniques.

Due to the increasing popularity of smartphones as VR devices, we intend to leverage the flexibility and scalability of gaze-based interaction in our future development efforts on these devices. We also foresee the integration of many more gaze-based actions in future applications. Some new actions under consideration include moving and destroying objects, as well as transport in the forms of teleportation and rappelling. Visual and usability enhancements to the current gaze marker systems are already in the project roadmap – the use of intelligent virtual guides for terrain navigation is being explored as the next major improvement.

Acknowledgment

The authors would like to acknowledge the following: Human Hands as Input Device for an Immersive Virtual Reality Experience (VREx) Project, Department of Science and Technology (DOST), the Philippine Council for Industry, Energy, Emerging Technology Research and Technology (PCIEERD), and Marish Viola for her contributions in game art.

References

[1] J. Pelz, M. Hayhoe, and R. Loeber, "The coordination of eye, head, and hand movements in a natural task," Experimental Brain Research, vol. 139, no. 3, 2001, pp. 266–277.

[2] C. Yu, P. Schermerhorn, and M. Scheutz, "Adaptive eye gaze patterns in interactions with human and artificial agents," ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 1, no. 2, 2012, p. 13.

[3] I. E. Sutherland, "The ultimate display," Multimedia: From Wagner to virtual reality, 1965.

[4] T. A. Furness, "The super cockpit and its human factors challenges," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 30, no. 1. SAGE Publications, 1986, pp. 48–52.

[5] "Oculus Rift," 2014, URL: https://www.oculus.com/ [accessed: 2015-11-24].

[6] "The best VR headsets 2015," 2015, URL: http://www.techradar.com/news/wearables/the-best-vr-headsets-2015-1292087 [accessed: 2015-11-24].

[7] W. Barfield, Fundamentals of wearable computers and augmented reality. CRC Press, 2015.

[8] K. Kato and H. Miyashita, "Creating a mobile head-mounted display with proprietary controllers for interactive virtual reality content," in Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology. ACM, 2015, pp. 35–36.

[9] "VR Best Practices Guidelines," 12 June 2015, URL: https://developer.leapmotion.com/vr-best-practices [accessed: 2015-11-27].

[10] N. Valdez, R. Besas, D. Dumalaon, and R. Atienza, "3d gestures on 2d screens for mobile games," in IEEE Asia Pacific Conference on Wireless and Mobile, 2014.

[11] R. C. Zeleznik, A. S. Forsberg, and J. P. Schulze, "Look-that-there: Exploiting gaze in virtual reality interactions," Technical Report CS-05, Tech. Rep., 2005.

[12] D. Charles, K. Pedlow, S. McDonough, K. Shek, and T. Charles, "Close range depth sensing cameras for virtual reality based hand rehabilitation," Journal of Assistive Technologies, vol. 8, no. 3, 2014, pp. 138–149. [Online]. Available: http://dx.doi.org/10.1108/JAT-02-2014-0007

[13] J. P. Costella, "A beginners guide to the human field of view," School of Physics, The University of Melbourne, 1995.

[14] T. Gowases, R. Bednarik, and M. Tukiainen, "Gaze vs. mouse in games: The effects on user experience," 2008.

[15] J. Schell, The Art of Game Design: A book of lenses. CRC Press, 2014.

[16] K. Jørgensen, "im overburdened!an empirical study of the player, the avatar, and the gameworld," in Proceedings from DiGRA. Citeseer, 2009.

[17] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster, "A dose of reality: Overcoming usability challenges in vr head-mounted displays," 2015.