Look to Go: An Empirical Evaluation of Eye-Based Travel in Virtual Reality

YuanYuan Qian
Carleton University
Ottawa, ON
Canada
heather.qian@carleton.ca

Robert J. Teather
Carleton University
Ottawa, ON
Canada
rob.teather@carleton.ca

ABSTRACT

We present two experiments evaluating the effectiveness of the eye as a controller for travel in virtual reality (VR). We used the FOVE head-mounted display (HMD), which includes an eye tracker. The first experiment compared seven different travel techniques to control movement direction while flying through target rings. The second experiment involved travel on a terrain: moving to waypoints while avoiding obstacles with three travel techniques. Results of the first experiment indicate that performance of the eye tracker with head-tracking was close to head motion alone, and better than eye-tracking alone. The second experiment revealed that completion times of all three techniques were very close. Overall, eye-based travel suffered from calibration issues and yielded much higher cybersickness than head-based approaches.

CCS CONCEPTS

• **Human-centered computing → Virtual Reality • Human-centered computing → Pointing Devices**

KEYWORDS

Travel performance, navigation, eye-tracking, head-mounted display, joystick, cybersickness.

1 INTRODUCTION

Eye-tracking is widely used in the HCI domain, commonly to augment usability evaluations, but also sometimes for eye-based interaction. Recent hardware advances have yielded low-cost head-mounted displays that include integrated eye trackers, such as the FOVE[1]. Other manufacturers such as Oculus[2] and Pupil Labs[3] have either just released or will soon release their own eye-tracking solutions. One of the more common applications of eye tracking in VR is foveated rendering [17,18], that can enhance immersion and user experience. Several studies [3,4,14,23] have investigated other applications of eye-tracking in VR. Eye-based interaction in VR – i.e., using the eye as an input controller – is comparatively understudied. A potential benefit of eye-based interaction in VR is that it may require less physical effort compared to other input devices (e.g., wands) to control the user viewpoint or manipulate objects, especially in large three-dimensional spaces [9].

According to Bowman et al.'s classic taxonomy [1], fundamental VR tasks include selection, manipulation, navigation, system control, and symbolic input. Navigation further breaks down into travel (moving oneself through a virtual environment) and wayfinding (the cognitive task of route planning through the virtual environment). Travel is a particularly interesting candidate for eye-based interaction. For example, Stellmach and Dachselt [24] conducted a study on VR travel using eye-based input. Their approach was indirect; participants used their tracked eyes to target 2D UI elements on a 2D panel to indirectly control the movement direction. We instead proposed to use the eye as a direct input control for travel via a modified gaze-directed steering. Effectively, this allows users to look where they want to go.

Gaze-directed steering (travel in the direction the head is looking) is a well-known travel metaphor [1,26,27] that has long been used for locomotion in VR, ever since first being proposed by Mine [15]. Variations are still common today in games such as *End Space VR*[4]. Looking in the direction we move is quite natural; eye tracking offers a more fine-grained approach to this that decouples movement from the head orientation, potentially allowing more natural interaction. Standard gaze-directed steering couples the view and movement directions, yet allows users to perform virtual walking or flying tasks at a fixed velocity fairly easily. Thus, we are interested in the potential advantages of combining head and eye-based input to leverage both benefits.

Our previous study [19] revealed that eye-tracking offers poor performance in 3D selection tasks. However, a few researchers [1,24] have considered the use of eye-tracking for travel. In general, travel has a lower accuracy requirement than selection. Travel techniques usually work well enough if users get to the general vicinity of where they intended to go. In light of this lowered accuracy requirement, it is reasonable that eye-based input may work better for travel than for selection. Thus, we developed two travel testbed virtual environments and conducted two experiments comparing the performance of eye-tracking as an alternative to other travel control techniques.

In the first study, participants flew through rings to compare gaze-directed steering using the eye to that with the head. For a

[1] www.getfove.com
[2] https://techcrunch.com/2016/12/28/the-eye-tribe-oculus/
[3] https://pupil-labs.com/vr-ar/
[4] http://endspacevr.com/

SUI '18, October 13–14, 2018, Berlin, Germany

ACM ISBN 978-1-4503-5708-1/18/10…$15.00
https://doi.org/10.1145/3267782.3267798

baseline, we also included mouse and joystick-based steering. This experiment included seven travel techniques in total: four "single input" and three "combination input". The single input techniques controlled both the head orientation and movement direction simultaneously, similar to first-person shooter game controls. These included: 1) head-only, 2) eye-only, 3) mouse-only, 4) joystick-only. With the exception of head-only, head-tracking was disabled in these input techniques. In the combination input methods, head-tracking was enabled in tandem with three other input techniques to control the travel direction. They were: 5) head+eye, 6) head+mouse, 7) head+joystick. We note that while mouse-based steering is atypical in VR travel, it is very common in first-person shooter games (used with keys to control movement speed). We included the joystick as another common representative of both game and VR travel. The head-tracking conditions were intended to isolate head-based input from eye-based input.

In the second experiment, participants followed a path along a terrain, walking to target cubes while avoiding obstacles. This experiment included only 3 travel techniques: 1) head-only, 2) eye-only, 3) joystick-only. The travel techniques controlled both the head orientation and movement direction simultaneously, as is typical in gaze-directed steering. Head-tracking was disabled in software for eye-only and joystick-only; this was a conscious decision to isolate eye-based control from head-based control. After all, if both were enabled, participants may simply use their head orientation. Other movement directions (e.g., side-to-side) were provided using the directional controls on an Xbox controller.

Our hypotheses included:

H1: Of the single input techniques, head-only would yield the lowest cybersickness due to providing consistent visual and vestibular information (for rotations).

H2: Of the single input techniques, eye-only would perform better than head-only and joystick-only since it both reduces the need for head motion, and due to the speed of saccades offering faster turning rates.

H3: Of the combination input techniques, participants will prefer head+eye since the other two combination input techniques need extra controller movements (operating mouse and joystick).

H4: All combination input techniques will offer better performance than their corresponding single techniques, which will be hindered by the absence of head-tracking.

Our main contribution is the first empirical study of the performance of the eye as a *direct* control device for travel in VR. A secondary contribution is the comparison of head-based travel and mouse/joystick-based travel, all common VR travel techniques.

2 RELATED WORK

2.1 Eye Movement Theory

The eyes utilize voluntary and involuntary movement to help acquisition, fixation and tracking. The brain exerts signals through three cranial nerves to control the six extraocular muscles which are attached to the eyeball, thus to control the eye movements [20]. The eyes never stop moving completely, even when fixated at one point. They are always making fast virtually random jittering movements. The photoreceptors and the ganglion cells cannot respond when a constant visual stimulus falls on them. In order to make the image received clearer, the random eye movement keeps changing the stimuli thus makes photoreceptors and the ganglion cells being active [21]. These short and rapid movements that occur when the eyes are scanning an area are referred to as saccades. The eyes move as fast as they can during a saccade, but the speed is not consciously controlled. It is useful to scan an area with the fovea of the eye in a high resolution [5]. The fovea is a small area of the retina of about 1° in size.

When watching a moving object or pursuing it, the head also moves to assist in tracking. But the head movement alone cannot catch up with fast moving objects [25]. In order to see the moving object clearly, the eyes move as well and try to keep the object image on the fovea. Lanman et al. [12] compared head and eye movement on trained monkeys when tracking moving objects. They pointed out that although eyes had irregular movement, the combination of head and eye could yield precise and smooth target pursuit contributed by the vestibular system. These results motivate the design of our head+eye travel techniques, which should perform better than the eye-only and head-only navigation.

2.2 Navigation in VR

There were a few relevant studies on travel test environments, techniques and evaluations that we used for inspiration. Nelson et al. [16] conducted a virtual flying study to evaluate a brain-body-actuated controller. They had two tasks: the first was to fly through hoops and as close to the center of the hoops as possible. The second was that there were ribbons connected between the hoops, they should fly within the boundaries of the hoops. Their post-test questionnaires were NASA task load index (TLX) and modified simulator sickness questionnaire (SSQ) [10]. We modelled our first travel task after this, and we employed similar metrics.

Cybersickness is similar to the motion sickness symptoms during or after exposure in a virtual environment [13]. Conflicts between the visual, vestibular, and proprioceptive senses are thought to yield cybersickness [11,13]. Thus, cybersickness is likely to occur when using the eyes or head alone to travel in VR. Hettinger et al. [7] indicated that a fixed-based visual display produced vection and sickness. When there is a significant mismatch between visual information and vestibular information (as is usually the case in VR travel supported by joysticks), people tend to experience motion sickness. Therefore, we expected that head-only would yield lowest levels of cybersickness.

Finally, Chen et al. [2] compared head- and joystick-based travel. They concluded that the head-based paradigm was superior to the joystick on user performance, presence and cyber sickness. We thus expect that our joystick conditions would offer lower performance and user satisfaction than head-related techniques.

2.3 Eye-Tracking in VR

Many researchers have noted the possibilities of using eye-tracking in VR. Several studies [3,4,14,23] employed eye tracking for applications other than interaction tasks.

To our knowledge, there are relatively few studies investigating performance of eye-based interaction in VR. Our previous study [19] compared performance of the eye and head in a 3D *selection* task. Head-only input offered better 3D selection performance than either eye-only input, or the combination of eye and head input. Similar results are reported by Hansen et al.[6]. We also look to studies of eye tracking in 3D games, which in some ways, are similar to VR. Isokoski and Martin [8] evaluated the effectiveness of eye-tracking to control aiming in a first-person shooter game as an alternative to the traditional mouse+keyboard. Smith and Graham [22] explored the eye-tracker as a control device in several video games, i.e., a first-person shooting game, a role playing game and an action/arcade game. Notably, they utilized eye-tracking to control view orientation in the FPS game, similar to our eye-only travel technique. They reported that although the eye performed slower than the mouse, the intuitive interactive way of eye-tracking increased immersion and significantly enhanced game experience.

Likely the closest study to ours is that of Stellmach and Dachselt [24], who also investigated eye-based input for virtual travel. The task involved navigating to a target position in 5 different difficulty levels. To complete the task, participants had to use their eyes to perform rotations and translations by looking at a 2D UI. They found that the continuous gradient-based input offered the fastest completion time and was most preferred by participants. Their post-test questionnaire employed Bowman's traveling questionnaire [1], which we also use in our experiments. The main difference between their study and ours is that our eye-based input operates by allowing participants to move in the direction they are looking; no additional UI elements are used. We argue that this is a potentially more natural approach.

3 EXPERIMENT 1

In this study, the task involved flying through rings in the air using seven different input techniques. We opted to start with a flying task since although 3D flying is potentially more complex and difficult than travel constrained to a terrain, it may also generalize to other travel tasks. It also applies in specific domains like gaming and flight simulation.

3.1 METHODOLOGY

3.1.1 Participants

We recruited fourteen participants (aged 18 to 40, μ = 27 years, 8 male). All were daily computer users (μ = 5 hours/day). Five had prior experience with eye tracking. Three had no prior VR experience, another three had limited VR experience (having used it once or twice ever), and the rest used VR on average around 5 times per month. All participants had colour vision. Five had normal vision, while the rest had corrected vision. All participants could see stereo, as assessed by pre-test trials. All participants were very familiar with games, 4 were frequently video game users (μ=5 times/week). One potential participant could not pass the calibration and two potential participants withdrew after the pre-test trials due to nausea.

3.1.2 Apparatus

The study was conducted using a VR-capable laptop with an Intel Core i7-7700HQ CPU, an NVIDIA GeForce GTX 1070 GPU, and 16GB RAM. Participants wore a FOVE VR HMD. The FOVE has a display resolution of 2560 x 1440 with a 100° field of view. It offers IMU-based sensing of head orientation, and optical tracking of head position, but does not provide interpupillary distance (IPD) correction. The FOVE includes two integrated infrared eye-trackers that offer tracking precision of less than 1° at a 120 Hz sampling rate. We also used a wired mouse and an Xbox controller.

We developed the software in Unity 5.5. The task involved flying through rings; to this end, the software presented three sets of rings in the air with the simple background of the blue sky over a desert and lake terrain. Participants were tasked with flying through these rings using the current control scheme. See Figure 1. The desert terrain was the reference object that enabled participants to feel the relative speed of motion. All the tasks were conducted in the air, no collisions occurred with the terrain.

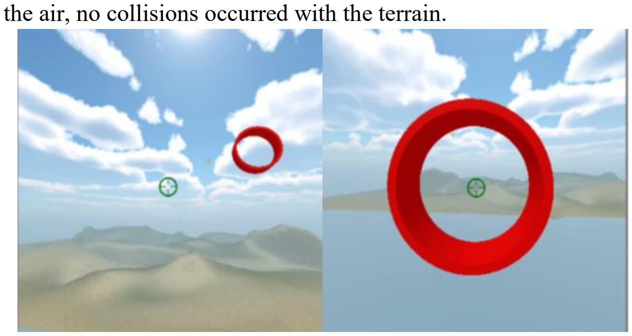

Figure 1. Experimental task showing the terrain, skybox, and rings the participants flew through.

The software presented eight yellow rings in a spiral arrangement. The spiral arrangement ensured that the participant had to control movement in eight directions during the trial. See Figure 2. The target ring was highlighted red. See Figure 1. Depending on the condition, the rings were put in 10º, 20º, or 30º deviations with respect to the previously passed ring. The distance (z-axis) between each ring was all 100 meters. The radius of each ring was 1.5 meters. The width of each ring was 1 m. The 1 m width ensured the software could reliably detect the collision point (in the plane of the ring) when the participant passed through the ring.

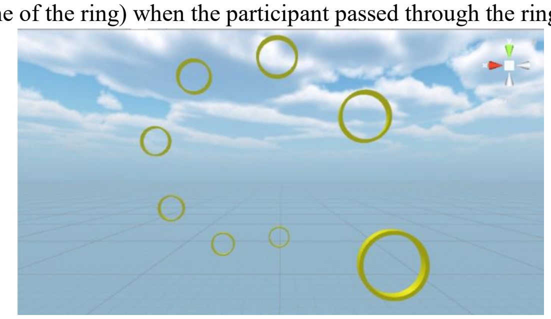

Figure 2. The ring arrangement, showing rings in 20° deviations, the middle difficulty level.

The frame rate was stable at 80 fps. To reduce cybersickness, we used a fixed velocity. We tested several velocities in the pilot study and finally chose the Unity default value, which seemed to yield lower sickness when tilting and rotating. In the non-joystick conditions, pressing the left mouse button started movement forward along the view vector. In the joystick-based conditions, we instead used the "A" button on the Xbox controller.

The software also displayed a green cursor to facilitate steering towards the targets (see Figure 1). The travel direction was controlled by moving this cursor in the view plane. In the four single input methods, this cursor was used to define the direction of movement vector, which originated at the camera in Unity. Lateral movement was not possible, and all movement was forward along the view direction. The cursor position was controlled using the following four single input methods:

Eye-only: Used the FOVE eye tracker. Gazing at a particular point would set the cursor to that position, rotating the viewpoint in that direction, and giving 1:1 control. The software continuously calculated the camera rotation angle using the eye ray provided by the FOVE on every frame.

Mouse-only: Used a desktop mouse to rotate the viewpoint. The cursor was fixed in the centre of the screen. This condition was very similar to first-person shooter games.

Joystick-only: Used the two axes input of one Xbox controller to rotate the viewpoint, similar to how the viewpoint is controlled in games controlled with joysticks. The "A" button on the controller activated forward movement.

Head-only: Used the FOVE's head-tracker to control the view direction. The cursor was fixed in the centre of the screen.

The three combination input methods (*head+mouse, head+eye, head+joystick*) operated similarly, except with the addition of head-tracking to control the camera's rotation. The other input controlled the cursor direction within the camera's view. As a result, the steering movement was the combined effect of both head movement and eye/mouse/joystick movement. The green cursor moved in the plane instead of being fixed in the centre of the screen. Table 1 summarizes the DOFs required with all input methods.

		DOF	Eye-only	Mouse-only	Joystick-only	Head-only	Head+mouse	Head+eye	Head+joystick
3D	2D	x*	②	②	②				
		y							
		Θz				◌	◌	◌	◌
		z*	●	●	●	●	●	●	●
		θx	●	●	●	●	●	●	●
		θy	●	●	●	●	●	●	●

Table 1. Degrees of freedom for each input method. Asterisk (*) indicates a DOF that used a separate key. Dashed-circles show supported but impractical DOFs (roll). Circles with 2 indicate a DOF that was only used in Exp. 2. All input methods were used in Exp. 1. Shaded input methods were the only ones used in Exp. 2.

Our study focused exclusively on steering effectiveness we did not support up/down or left/right translations – only forward motion (along the view vector, as described above). Thus, all single input techniques supported 3DOF input: yaw (θy), pitch (θx), and z translation by pressing the corresponding button. The combination techniques added a single DOF, roll (θz). However, because of the nature of the task, head roll was not really practical and is set in light grey in Table 1.

The software recorded all the coordinates of the collision points with the plane of each ring (to facilitate accuracy measures, i.e., distance from the ring centre), including inside and outside the ring, the successes and failures.

3.1.3 Procedure

Upon arrival, we briefed participants on the motivation, goals, and procedure for the experiment, then provided them with consent forms and demographics questionnaires. Participants then viewed a demo video of the interface and were introduced on how to operate each travel technique. All participants first completed the FOVE calibration process, which took approximately one minute. Calibration involved gazing at a green dot that appeared at a circular position on the display. We also used this calibration process as pre-screening for the participants: Potential participants who could not complete the calibration process could not take part in the experiment. Prior to each new session using the eye tracker (i.e., eye-only and head+eye), the eye tracker was re-calibrated to ensure accuracy. Since all participants had prior experience with the mouse and joysticks, and many (8/14) participants indicated that they were very familiar with the use of head-based orientation in VR, we added a few practice trials for the unfamiliar travel techniques that used the eye-tracker (head+eye and eye-only).

Participants were instructed to fly through the red highlighted ring, and as close to the centre as possible. Upon commencing testing, all of rings appeared in front of the view and the first target ring was red. Due to the distance between rings, participants were not able to initially see all rings, but could see the next three or four rings in the view. As participants travelled through the rings the remaining rings appeared. Upon passing each red (target) ring, it would disappear and the next ring in the sequence would turn red.

A block involved passing through 8 rings, each representing a different trial, and each in one of 8 different directions, organized in a spiral/corkscrew configuration. See Figure 2. Each travel technique testing session consisted of 3 such blocks. An extra "practice" ring was added to each block to help participants get used to a new condition. Data for this practice ring was excluded from our analysis. Regardless if the participant flew through, or missed (outside) the target ring, the next ring would highlight red. If they flew outside the ring, the trial was recorded as a miss.

Upon completing a session, participants completed three questionnaires, the NASA-TLX, the SSQ, and the traveling performance questionnaire developed by Bowman et al. [1]. Finally, we also debriefed participants in a short interview. Our experiment took approximately 70 minutes in total for each participant, for which they were compensated $10.

3.1.4 Design

The experiment employed a 7 × 3 within-subjects design. The independent variables and their levels were as follows:

Travel technique: Eye-only, head-only, mouse-only, joystick-only, head+eye, head+mouse, head+joystick

Difficulty: 10°, 20°, 30°

Since we considered each ring a single trial, in total, each participated completed 7 × 3 × 8 × 3 = 504 trials. Across all 14 participants, this yielded 7056 trials. Difficulty was represented as eccentricity of the next ring (i.e., necessitating a 10°, 20°, or 30° rotation of the viewpoint from the previous ring). Difficulty was arranged from the easiest to the hardest, i.e., the first three blocks were 10°deviations, the second three blocks were 20° deviations, the last three blocks were 30° deviations. Ordering of travel technique was counterbalanced according to a Latin square.

The dependent variables were completion time, success rate, collision radius, NASA-TLX and SSQ. Completion time was average time to complete one trial. Success rate was the percentage of rings successfully passed per difficulty in each session (i.e., percentage of rings *not* missed). Collision radius represented the mean distance from the centre of the ring.

3.2 RESULTS AND ANALYSIS

3.2.1 Completion time

Mean completion times per trial are summarized across the travel techniques and three difficulty levels in Figure 3. There was a significant main effect of travel technique on the completion time ($F_{6,273} = 38.607$, $p < .001$), but the main effect for difficulty on the completion time was not significant ($F_{2,273} = 0.96$, ns). The interaction effect was also not significant ($F_{12,273} = 0.248$, ns).

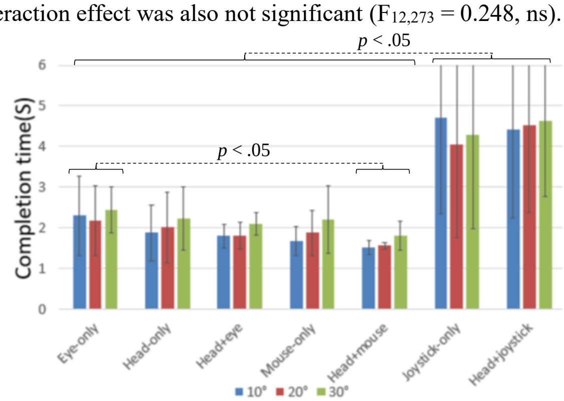

Figure 3. Mean completion time by travel techniques on three difficulty levels. Error bars show ±1 SD. Braces and dashed lines indicate "clusters" of travel techniques that show pairwise significant differences via post-hoc testing (*p* < .05).

Overall, participants tended not to take much longer regardless of difficulty. The reason might be that degree deviations were insufficient to truly create notably different difficulty levels. A Tukey-Kramer post-hoc revealed significant pair-wise differences between some of the travel techniques. Notably, both joystick techniques yielded much worse completion times than all other travel techniques. The use of head-tracking did not help improve the speed. Additionally, the head+mouse travel technique was significantly faster than eye-only. The rest of the travel techniques were not significantly different from one another. Pairwise differences are summarized in Figure 3.

3.2.2 Success rates

Figure 4 depicts success rate by difficulty level for each travel technique. There was a significant main effect of travel technique on success rate ($F_{6,273} = 20.41$, $p < .001$). Neither the main effect for difficulty level was significant ($F_{2,273} = 1.449$, $p > .05$), nor was the interaction effect ($F_{12,273} = 0.232$, ns). A Tukey-Kramer post-hoc test (seen in Figure 4) revealed pair-wise differences ($p < .05$).

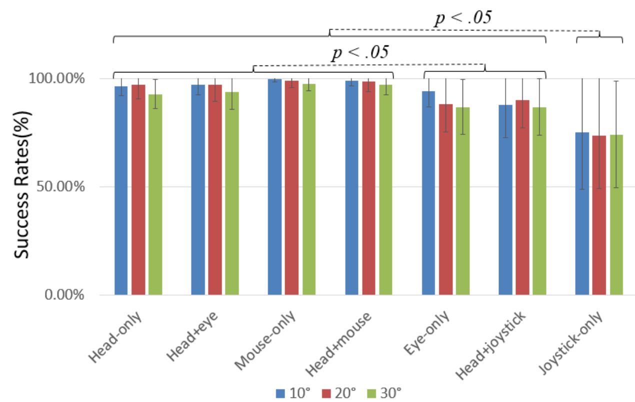

Figure 4. Mean success rate by travel techniques and difficulty level. Error bars show ±1 *SD*. Braces and dashed lines indicate "clusters" of travel techniques that show pairwise significant differences via post-hoc testing (*p* < .05).

3.2.3 Coordinate Map and Collision Radius

Figure 5 shows coordinate maps for each travel techniques cut from the *z*-axis plane of all collisions within a 3 m radius of the ring. The red circle shows the target ring with 1.5 m radius.

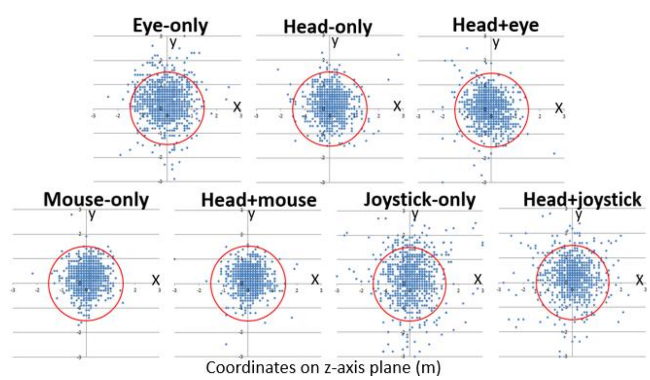

Figure 5. Coordinate maps on z-axis plane for each travel technique, across all three difficulty levels. The red ring depicts the target ring, and each blue mark depicts a coordinate. This includes all trials for each travel technique, aggregated together.

This visualization gives a good indication of the degree of control offered by each of the travel techniques; conditions more closely clustered near the centre of the red circle indicate conditions where participants were better able to stay near the ring centre while traveling. Conversely, conditions with many data points outside the circle indicate travel techniques where participants had greater difficulty. Mouse-only offered consistently high precision, hitting virtually all the collisions within the ring. Head+eye was a bit sparser than head-only, but both did well overall. Eye-only, joystick-only and head+joystick all had many collisions out of the ring, while joystick-only was the worst. This map revealed the consistency with success rates.

We also analyzed the mean collision radius – i.e., the magnitude of error from the target centre. These scores are seen in Figure 6. The radius represents how far the actual path deviated from the optimal path. Since participants were instructed to try to hit the centre of a ring, the greater the radius, the less accurate the travel technique was. There was a significant main effect of travel technique on the collision radius ($F_{6,273} = 15.108$, $p < .001$), but no significant effect for difficulty level ($F_{2,273} = 2.192$, $p > .05$) nor the interaction effect ($F_{12,273} = 0.09$, ns). The Tukey-Kramer post-hoc test showed pair-wise differences ($p < .05$) between several travel techniques, as summarized in Figure 6.

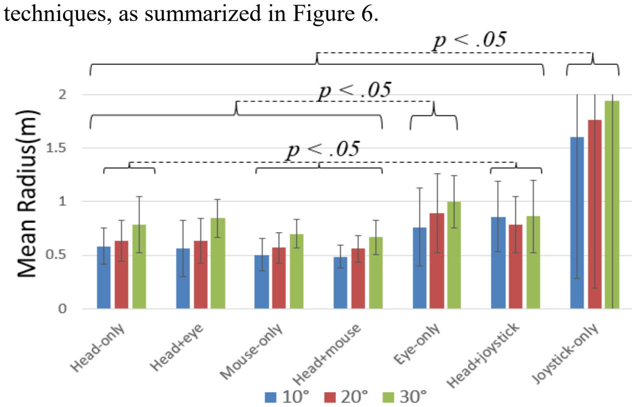

Figure 6. Mean radius of the collision points of 10, 20 and 30-degree levels. Error bars show ±1 *SD*. Braces and dashed lines indicate "clusters" of travel techniques that show pairwise significant differences via post-hoc testing (*p* < .05).

3.2.4 Subjective Measures

We included three questionnaires to garner subjective data on the conditions. The first was the 5-item travel performance questionnaire and based on Bowman's travel questionnaire [1]. We asked participants to fill this questionnaire after finishing each condition. Each participant rated perceived speed, accuracy, spatial awareness, ease of learning, and ease of use on a 5-point scale, with 5 as the most favourable score. Scores from this questionnaire are seen in Figure 7. Overall, participants rated mouse-only and head+mouse the best on all points. Head-only was rated lower than head+mouse, but still higher than eye-only and head+eye on all points. Head+eye was better than eye-only on spatial awareness, while eye-only is better on the learnability.

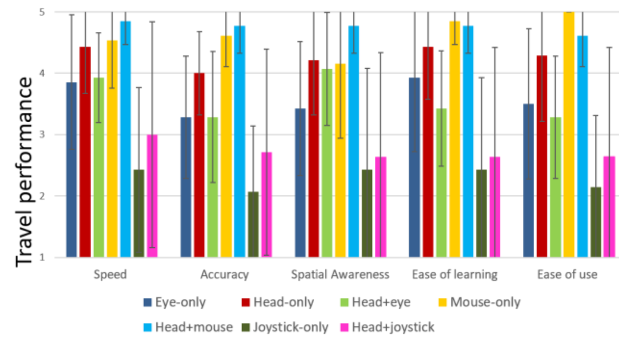

Figure 7. Average of response scores for travel performance question. Error bars show ±1 SD. Higher scores are more favorable in all cases.

We also conducted the simulator sickness questionnaire (SSQ), [10] to assess participant cybersickness levels. The questionnaire consists of 16 items with 3 weighted symptom categories, i.e., nausea, oculomotor, and disorientation. Participants completed the SSQ after finishing each condition. Joystick-only, eye-only and head+joystick had much higher symptoms than other techniques on all three profiles. The joystick-based techniques were worst, but eye-only also had high symptoms. In general, in the absence of head-tracking (e.g., eye-only), or in conditions with inconsistent visual-vestibular cues (the joystick-based conditions), participants experienced worse symptoms.

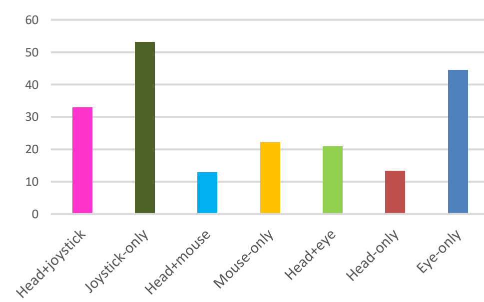

Figure 8. Total weighed scores for SSQ by travel technique.

Finally, we also used the NASA-TLX questionnaire to evaluate workload for each travel technique. Each response was rated on a 21-point scale, with 21 as the least favourable response and 1 the most favourable response for performance, vice versa for other 5 items. Scores are seen in Figure 9. Unsurprisingly, and consistent with our objective performance measures, mouse-only and head+mouse were rated the lowest on all scales, followed by head+eye, head+only and eye+only. The joystick techniques were rated the worst.

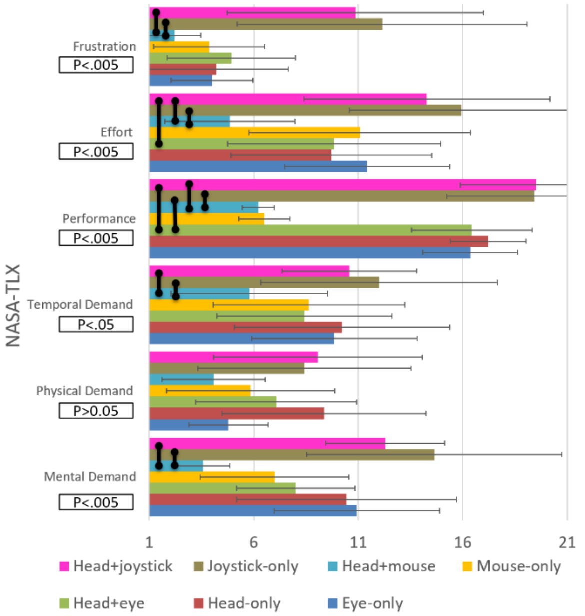

Figure 9. Average of response scores for each NASA-Task Load Index question. Error bars show ±1 SD. Higher scores are less favorable in all cases. Statistical results via the Friedman test shown to the left. Vertical bars (⟷) show pairwise significant difference.

4 EXPERIMENT 2

This experiment used terrain-constrained movement rather than flying and was expected to generalize better to more "realistic" travel tasks. After all, most VR environments employ such physical constraints. This experiment included only a subset of travel techniques from the first. We excluded the combination techniques to simplify the experiment design and to focus on the effectiveness of eye-based travel in isolation from the other conditions.

4.1 METHODOLOGY

4.1.1 Participants

We recruited twelve participants (aged 18 to 50, μ = 32 years, 9 male). All were daily computer users (μ = 5 hours/day). Three had limited prior experience with eye tracking (having used it once or twice ever). Four had no prior VR experience, five had limited VR experience (having used it once or twice ever), and the rest used VR around 5 times per month. All participants had normal or corrected colour vision. All participants could see stereo, as assessed by pre-test trials. All participants were very familiar with games, nine were frequently video game users (μ=5 times/week). Two potential participants could not pass the calibration.

4.1.2 Apparatus

We utilized the same VR-capable laptop and FOVE HMD as in the first study. We also used the directional pad (D-Pad) on an Xbox controller to provide lateral walking movement (side-to-side), in addition to forward and backwards movement. Viewpoint direction was controlled by either the eye tracker (eye-only), head orientation (head-only) or the joystick (joystick-only). The head-tracking was disabled on eye-only and joystick-only conditions. All operated as described in Experiment 1, with the exception that the software did not display the green cursor. The rotation angle of the input device controlled the movement vector orientation. That is, head-only used the head orientation. The eye controlled the movement vector by continuously calculating the rotation angle of the rays from the eye. The joystick's two axes input controlled the camera's rotation.

Table 1 shows the DOFs provided by each of the three input methods. Note that each supports one additional DOF compared to its Experiment 1 variant, due to the addition of side-to-side stepping motions via the Xbox directional pad.

We developed the experimental interface in Unity 5.5. The software presented three sets of waypoints – represented as grey boxes – following a path along a circular road around a lake. Participants were tasked with walking to the active waypoint (displayed in red) using the current control scheme. The task was alternately presented with and without obstacles, represented using tires positioned in the road. In the obstacles condition, the tires were positioned in the way between subsequent waypoints; participants had to avoid these. Bumping into obstacles was recorded, and moreover hindered participants' forward progress, yielding a worse time score. Figure 10 depicts trials both with and without obstacles. Figure 11 depicts an overview of the scene.

Figure 10. Experimental task showing the waypoints (red target blocks) both with and without tire obstacles.

Figure 11. The top-down view of the waypoints in the zig-zag pattern with tire obstacles.

The task presented ten waypoints in a zig-zag arrangement on the road. See Figure 11. Each waypoint was a 0.5 × 0.5 × 4 m box. The waypoints were randomly positioned between 20º to 30º deviations with respect to the previously reached waypoint. The distance between each waypoint was randomly chosen between 30 and 50 meters. We put one to three tires randomly in positions between the cubes as obstacles. The tires were 2 meters in diameter.

4.1.3 Procedure

Upon arrival, we first briefed participants on the experiment motivation and procedure, then provided consent forms and demographics questionnaires. Then we provided a demo video of the interface and introduced them how to operate each of the travel techniques. All participants first completed the FOVE calibration process, which took approximately one minute. We utilized the directional pad on an Xbox controller to control movement in all conditions. Viewpoint rotation (and hence movement orientation) was controlled by the active travel technique.

Participants were instructed to walk to the red box on the road as quickly as possible. If the participants collided with a tire, it would not disappear, and they would have to move around it to bypass it. These obstacles were intended to add some extra challenge to the task, as well as additional realism, since general travel tasks are rarely free of obstructions. Upon touching the red waypoint, it would become grey again, and the next waypoint would turn red (becoming active). Upon starting the experiment, all of waypoints appeared in front of the participant, and only the first was red. Due to the distance between waypoints, participants could not initially see all of them due to occlusion and perspective scaling. However, they could see the next three or four waypoints, and the rest came into view as they progressed along the path. A block consisted of 10 waypoints, and each session with a travel technique consisted of 3 such blocks. An extra "practice" waypoint was added to each block to help participants get used to a new condition. Data for this practice trial was excluded from our analysis.

Upon completing a session, participants completed three questionnaires, the NASA-TLX, the SSQ, and the travel questionnaire developed by Bowman et al. [1]. Finally, we also debriefed participants in a short interview.

4.1.4 Design

The experiment employed a 3 × 2 within-subjects design. The independent variables and their levels were as follows:

Travel technique: Eye-only, head-only, joystick-only

Obstacles: On, Off

Each waypoint was considered a single trial. In total, each participated completed 2 × 3 × 10 × 3 = 180 trials. Across all 12 participants, this yielded 2160 trials. Ordering of travel technique and obstacles was counterbalanced according to a Latin square.

The dependent variables were completion time, travel performance, NASA-TLX and SSQ. We also used a pathfinding algorithm to get the shortest possible time for each condition, and provide this as a baseline (i.e., best possible performance achievable) as comparison with the other travel techniques.

4.2 RESULTS AND ANALYSIS

4.2.1 Completion time

Mean completion times are summarized across the travel techniques for both conditions with and without obstacles in Figure 12. The mean completion time were the total completion time per session per participant. We did not compare across obstacles-on and obstacles-off conditions, since the presence of obstacles changed the task sufficiently to invalidate such a comparison.

For trials with obstacles, there was a significant main effect of travel technique on completion time ($F_{2,22} = 3.336$, $p < .05$). The completion time that participants took were also very close. The eye-only and joystick-only took slightly longer time than the head-only. The Tukey-Kramer post-hoc test showed pair-wise differences ($p < .05$) between travel techniques as depicted in Figure 12.

For obstacles off trials, there was a significant main effect of travel technique on the completion time ($F_{2,22} = 3.415$, $p < .05$). Overall, the completion time that participants took were very close. The eye-only took slightly longer time than the others, while the joystick-only was slightly longer than head-only. Although we used the Tukey-Kramer post-hoc test, it failed to detect pair-wise significant differences.

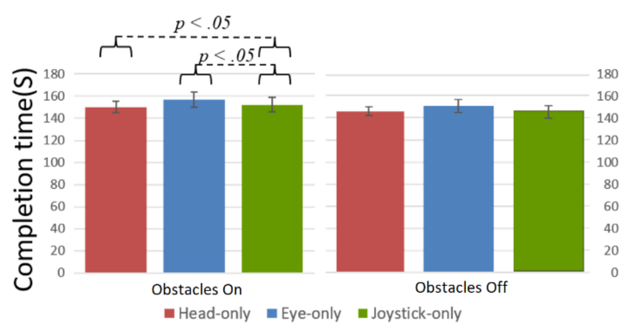

Figure 12. Mean completion time by travel techniques without obstacles. Error bars show ±1 SD. Braces and dashed lines indicate "clusters" of travel techniques that show pairwise significant differences via post-hoc testing at the *p* < .05 level

4.2.4 Subjective Measures

We included three questionnaires to garner subjective data on the conditions. The first was the travel performance questionnaire consisting of 5 items, and based on Bowman's travel questionnaire [1]. We asked participants to fill this questionnaire after finishing each travel technique. The score included overall rates on two difficulty levels. Each participant rated on a 5-point scale, with 5 as the most favorable response and 1 the least favorable response.

Scores from this questionnaire are summarized in Figure 13. Overall, participants rated head-only the best on all points. Eye-only was better than joystick-only on speed, accuracy and spatial awareness. However, eye-only did not perform well on the learnability and usability compared by joysticks. This is likely because participants generally had some prior experience with joysticks.

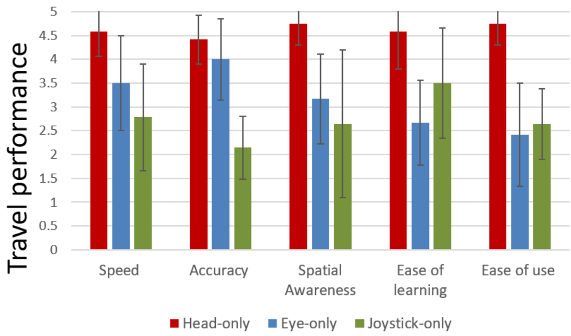

Figure 13. Average of response scores for travel performance question. Error bars show ±1 SD. Higher scores are more favorable in all cases.

We conducted the simulator sickness questionnaire (SSQ), based on Kennedy et al. [10]. We asked participants to fill this questionnaire after finishing each input technique. Eye-only had the highest sickness symptoms, followed by joystick-only.

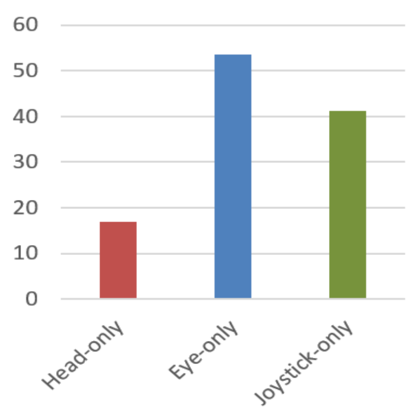

Figure 14. Total weighed scores for SSQ question

Finally, we also used the NASA-TLX questionnaire to evaluate the workload for each travel technique. Each response was rated on a 21-point scale, with 21 as the least favourable response and 1 the most favourable response for performance, vice versa for other 5 items. Scores are seen in Figure 15. Head-only had the best scores on every point expect for physical demand. Eye-only was better on effort and mental demand than joystick-only.

5 Discussion

After the experiment, we conducted short interviews with participants. We asked them their preference towards each travel technique. In the first study, most participants liked head+mouse the most, while in the second study, they liked the head-only technique the most. Participants generally felt the most comfortable and confident when using head orientation to control their travel direction, across both studies. However, the eye-based techniques

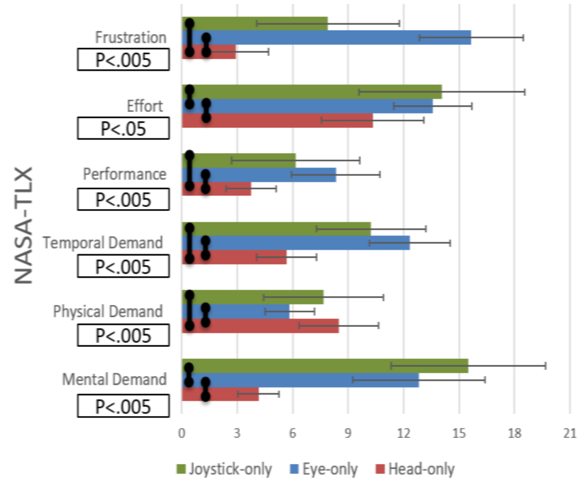

Figure 15. Average of response scores for each NASA-Task Load Index question. Error bars show ±1 SD. Higher scores are less favorable in all cases. Statistical results via the Friedman test shown to the left. Vertical bars (⟜⟞) show pairwise significant difference.

were also mentioned favourably by participants. In the first study, five participants rated the head+eye and three rated the eye-only as the second favourite technique. They did not select head-only because it caused much more movement than the eyes. In the second study, the participants tended to prefer the joystick over the eye, perhaps because of extensive experience with joysticks.

In terms of eye-based techniques, calibration and learning effects influenced the performance in both experiments. Most participants had never used eye-tracking in VR, so all experienced some degree of learning and adaptation depending on individual differences. As mentioned earlier, in anticipation of this, we had added a few extra practice trials for the eye-based techniques. However, in practice, these extra trials were likely insufficient to level the playing field. Most participants adapted to eye-control in around a minute of practice, but some took slightly longer. However, we found the extra few minutes' trials would not be sufficient for this novel technique, suggesting the need for a future longer-term study. A few participants also commented on this, suggesting that more training would help eye-based performance. Unfortunately, because in the limitation of the entire experiment time, we did not provide them more training trials than 4 minutes, including each time for calibration. However, these conclusions were all based on participants' subjective perspectives and our observations. Thus, we expect that a longitudinal study would reveal more realistic results on the long-term potential for eye-based travel control in VR.

Notably, we experienced many calibration issues, which further limited the potential of the eye-based techniques. A few potential participants could not pass the calibration after more than 5 attempts. Two potential participants passed the calibration but could still not control their eyes properly, i.e., they lost the orientation after calibration and could not focus on the target ring using their eyes. We tried to recalibrate five times but they still could not control the cursor. This yielded a great degree of jitter, which in turn caused a moderate level of cybersickness. We thus stopped the trials for these participants and they withdrew from the experiment.

Other participants also felt a certain level of nausea in the first few trials or in the middle of the session when inaccuracy occurred. This likely contributed to the higher SSQ levels with the eye-only travel technique. Cybersickness was likely also influenced by the absence of head-tracking in some conditions; this introduces another visual-vestibular conflict.

During the experiment, we found that if the participants did not tie the HMD belt very tightly, the relative distance would be changed after moving the head yielding inaccuracy. Most of participants could notice the accuracy decreasing rapidly after a few trials. When this happened, we asked them to recalibrate the eye tracker and restart the session. The combination of the calibration mechanism, HMD weight, and the design of headband all influenced the accuracy. In the head+eye session of first study, the head likely compensated for the limits of eye calibration, the participants could adjust the move direction by moving their head slightly as long as the movement was not so strenuous to change the relative position between the HMD and eyes.

On a more promising note, many of these issues are likely due to hardware limitations of the FOVE eye-tracker and could be potentially addressed with better and/or more expensive eye-tracking hardware. In this sense, it is exciting that there is interest in eye tracking among many HMD manufacturers – it seems likely that better hardware will become available soon. Despite these limitations, and as noted earlier, some participants still felt favourably towards the eye tracker conditions, and performance results were not substantially worse (especially in Experiment 2). We are thus somewhat optimistic about these results.

Overall, joystick-only performed the worst across all dependent variables in the first study, but better than eye-only in the second study. Four participants with extensive gaming experience found the joystick quite natural and comfortable, but they pointed out it was always harder to control the joystick in the air than on the ground, which would be the reason that the joystick performed better on a terrain than in the air. In the first study, the head+joystick had higher standard deviations than others for completion time and success rates. The reason might be the different traveling strategies used by participants. Some participants liked to use the joystick as the dominant technique but a few of them liked to use the head as the dominant technique especially for larger degree deviations between rings. Thus, this was also the reason that we still employed joystick-only in the second study. The second study implied that the joystick could perform better in a casual task on a land, while the eye could perform better in an intensive task that needs quick response.

In reviewing our hypotheses, we confirmed that head-only yielded the least cyber sickness, head+eye performed better than head+joystick in the air, head+eye and head+joystick improved their corresponding single input techniques on for all objective evaluations and subjective feelings. However, eye-only did not perform better than head-only.

6 CONCLUSIONS

We developed two different testbeds for VR navigation. We implemented seven input techniques for a flying experiment and three input techniques for walking experiment. We explored the performance of eye-based travel techniques in VR based on our flying and walking testbeds. Results of the first study indicated that the completion time and success rates of head+eye were very close to head-only. The second study showed that the completion time of eye-only was a bit longer than head and joystick, but very close to head and joystick. However, calibration issues and learning effects noticeably influenced the eye-only input technique, which also yielded high cybersickness due to the absence of head tracking. In subjective questionnaires, the participants generally rated the head-based travel techniques higher than eye-based, while joystick-based were the worst in flying performance but better than the eye in walking performance. Notably, the participants rated head+eye higher than head-only and eye-only in NASA-TLX, that also confirmed that the combination of head and eye worked better and compensated the imprecision of the eye-tracker.

In both studies, we observed different learning rates; as expected, participants performed better with eye-tracking by the end of the sessions than the beginning, despite pre-test practice trials. This suggests a longitudinal study would be required to get a true sense of the comparative effectiveness of the controllers. Future work could investigate how long it takes for participants to adapt to eye-based interaction in VR. Future work would also focus on eye-based interaction in VR using a broader range of tasks (e.g., manipulation) and enhanced task realism (e.g., selecting targets outside the field of view).

Ultimately, although eye tracking did not perform better than head-based input, the results – both objective and subjective – were quite close. In the first experiment, eye-tracking even outperformed the much more familiar joystick. We speculate that simply using a better eye-tracker might make eye-based travel much more competitive. To this end, we continue to be optimistic about upcoming eye-tracking head-mounted displays.

ACKNOWLEDGEMENTS

Thanks to all participants. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. D.A. Bowman, D. Koller, and L.F. Hodges. 1997. Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, 45–52,. https://doi.org/10.1109/VRAIS.1997.583043
2. Weiya Chen, Anthony Plancoulaine, Nicolas Férey, Damien Touraine, Julien Nelson, and Patrick Bourdot. 2013. 6DoF Navigation in Virtual Worlds: Comparison of Joystick-based and Head-controlled Paradigms. *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology - VRST '13*: 111. https://doi.org/10.1145/2503713.2503754
3. Andrew T. Duchowski, Vinay Shivashankaraiah, Tim Rawls, Anand K. Gramopadhye, Brian J. Melloy, and Barbara Kanki. 2000. Binocular eye tracking in virtual reality for inspection training. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*: 89–96. https://doi.org/10.1145/355017.355031

4. Kai Essig, Daniel Dornbusch, Daniel Prinzhorn, Helge Ritter, Jonathan Maycock, and Thomas Schack. 2012. Automatic analysis of 3D gaze coordinates on scene objects using data from eye-tracking and motion-capture systems. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*, 37. https://doi.org/10.1145/2168556.2168561
5. John M. Findlay. 2009. Saccadic eye movement programming: Sensory and attentional factors. *Psychological Research* 73, 2: 127–135. https://doi.org/10.1007/s00426-008-0201-3
6. John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN '18*. https://doi.org/10.1145/3206343.3206344
7. L. J. Hettinger, K. S. Berbaum, R. S. Kennedy, W. P. Dunlap, and M. D. Nolan. 1990. Vection and simulator sickness. *Military psychology : the official journal of the Division of Military Psychology, American Psychological Association* 2, 3: 171–181. https://doi.org/10.1207/s15327876mp0203_4
8. Poika Isokoski and Benoît Martin. 2006. Eye tracker input in first person shooter games. *Proceedings of the 2nd Conference on Communication by Gaze Interaction: Communication by Gaze Interaction-COGAIN 2006: Gazing into the Future*: 4–6. Retrieved from http://www.sis.uta.fi/~pi52316/cogain2006/cogain2006.pdf
9. Robert J.K. Jacob and Keith S. Karn. 2003. Eye Tracking in Human-Computer Interaction and Usability Research. Ready to Deliver the Promises. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*: 531–553. https://doi.org/10.1016/B978-044451020-4/50031-1
10. Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3: 203–220. https://doi.org/10.1207/s15327108ijap0303_3
11. James R. Lackner. 1992. Simulator sickness. *The Journal of the Acoustical Society of America* 92, 4: 2458–2458. https://doi.org/10.1121/1.404501
12. J. Lanman, E. Bizzi, and J. Allum. 1978. The coordination of eye and head movement during smooth pursuit. *Brain Research* 153, 1: 39–53. https://doi.org/10.1016/0006-8993(78)91127-7
13. Joseph J. LaViola. 2000. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin* 32, 1: 47–56. https://doi.org/10.1145/333329.333344
14. Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based head gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*. https://doi.org/10.1145/2168556.2168578
15. M Mine. 1995. Virtual environment interaction techniques. *UNC Chapel Hill Computer Science Technical Report TR95-018*: 507242–507248.
16. W. Todd Nelson, Lawrence J. Hettinger, James A. Cunningham, Merry M. Roe, Michael W. Haas, and Leon B. Dennis. 1997. Navigating through virtual flight environments using brain-body-actuated control. *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*: 30–37. https://doi.org/10.1109/VRAIS.1997.583041
17. Yun Suen Pai, Benjamin Tag, Benjamin Outram, Noriyasu Vontin, Kazunori Sugiura, and Kai Kunze. 2016. GazeSim: Simulating Foveated Rendering Using Depth in Eye Gaze for VR. *ACM SIGGRAPH 2016 Posters on - SIGGRAPH '16*: 1–2. https://doi.org/10.1145/2945078.2945153
18. Anjul Patney, Joohwan Kim, Marco Salvi, Anton Kaplanyan, Chris Wyman, Nir Benty, Aaron Lefohn, and David Luebke. 2016. Perceptually-based foveated virtual reality. In *ACM SIGGRAPH 2016 Emerging Technologies on - SIGGRAPH '16*, 1–2. https://doi.org/10.1145/2929464.2929472
19. Yuan Yuan Qian and Robert J Teather. 2017. The eyes don't have it. In *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17*, 91–98. https://doi.org/10.1145/3131277.3132182
20. H. Ridley. 1973. Eye and Brain. *BMJ* 4, 5893: 682–682. https://doi.org/10.1136/bmj.4.5893.682-b
21. John P. Seward. 1942. Psychology: The science of behavior. *Psychological Bulletin* 39, 10: 887–889. https://doi.org/10.1037/h0052552
22. J. David Smith and T. C. Nicholas Graham. 2006. Use of eye movements for video game control. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology - ACE '06*, 20. https://doi.org/10.1145/1178823.1178847
23. India Starker and Richard A Bolt. 1990. A gaze-responsive self-disclosing display. *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*, April: 3–10. https://doi.org/10.1145/97243.97245
24. S. Stellmach and R. Dachselt. 2012. Designing gaze-based user interfaces for steering in virtual environments. *Proc. of ETRA'12* 1, 212: 131–138. https://doi.org/10.1145/2168556.2168577
25. Gerald Westheimer and Suzanne P McKee. 1975. Visual acuity in the presence of retinal-image motion. *Journal of the Optical Society of America* 65, 7: 847–850. https://doi.org/10.1364/JOSA.65.000847
26. Catherine Zanbaka, Benjamin Lok, Sabarish Babu, Dan Xiao, Amy Ulinski, and Larry F. Hodges. 2004. Effects of travel technique on cognition in virtual environments. In *Proceedings - Virtual Reality Annual International Symposium*, 149–156. https://doi.org/10.1109/VR.2004.1310068
27. Robert C Zeleznik, Andrew S Forsberg, and Jurgen P Schulze. 2005. Look-That-There: Exploiting Gaze in Virtual Reality Interactions. *Technical Report CS-05-04*.