The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection 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 a study comparing selection performance between three eye/head interaction techniques using the recently released FOVE head-mounted display (HMD). The FOVE offers an integrated eye tracker, which we use as an alternative to potentially fatiguing and uncomfortable head-based selection used with other commercial devices. Our experiment was modelled after the ISO 9241-9 reciprocal selection task, with targets presented at varying depths in a custom virtual environment. We compared eye-based selection, and head-based selection (i.e., gaze direction) in isolation, and a third condition which used both eye-tracking and head-tracking at once. Results indicate that eye-only selection offered the worst performance in terms of error rate, selection times, and throughput. Head-only selection offered significantly better performance.

CCS CONCEPTS

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

KEYWORDS

Selection performance, eye-tracking, head-mounted display, ISO 9241-9, Fitt's law

ACM Reference format:

Y. Qian, R. J. Teather. 2017. The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the ACM Symposium Spatial User Interaction, Brighton, UK, October 2017 (SUI'17)*, 8 pages. DOI: 10.1145/3131277.3132182

1 INTRODUCTION

Target selection, or target acquisition [26], is a critical user interface task, and involves identifying a specific object from all available objects. As early as 1984, Foley et al. [9] recognized the importance of target selection, and analyzed selection tasks for 2D GUIs. Since then, many researchers [20, 24, 26, 30] have investigated and evaluated 3D selection in virtual and augmented reality environments. Many innovative selection metaphors emerged, such as the virtual hand [23], ray-casting [20], and image plane interaction [22]. These interaction techniques are based on movement of the hand, or in some cases, the head. Modern virtual reality (VR) systems mostly continue this trend. Head-mounted displays (HMDs) that include a handheld tracked input device, such as the HTC *Vive*, or Oculus *Rift*, tend to use virtual hand or ray-based interaction. HMDs that do not include such an input device, such as the Microsoft *Hololens* and Samsung *Gear VR*, instead tend to necessitate the use of gaze direction (i.e., user head orientation) coupled with gestures (e.g., airtap) for interaction. These methods are imprecise, and may yield neck fatigue.

Eye-tracking offers a compelling alternative to head-based selection. Previous 2D selection research has revealed that eye-tracking can even offer comparable performance to the mouse, in certain cases [23]. This has only recently become a viable option in VR due to the advent of inexpensive eye-tracking HMDs such as the FOVE (see https://www.getfove.com). The FOVE is the first commercially available eye-tracking VR HMD. It enables the use of eye tracking as a selection technique; Users can control a cursor to select objects simply using their eyes. However, the performance of eye-based *selection* has not previously been studied in VR contexts. The motivation of our work is thus to compare the performance of both eye and head-based selection, both in isolation from one another, and in tandem.

We conducted an experiment based on the international standard, ISO 9241-9 [12], which utilizes Fitts' law [9] to evaluate pointing devices [34]. We compared three different selection techniques using the FOVE: 1) eye-based selection without head-tracking, which we dub *eye-only* selection, 2) head-based selection without eye-tracking, dubbed *head-only* selection, and 3) eye-tracking and head-tracking enabled at the same time, henceforth *eye+head* selection. We compared these selection techniques across several different combinations of target size and depth, based on past work in 3D selection [29]. Our hypotheses included:

H1: Eye+head would offer the best of speed among the three selection techniques, because humans are already well-adapted to coordinating eye and head movement [15].

H2: Head-only would offer the lowest error rate, due to the inherent imprecision of eye-tracking.

H3: Eye-only selection would be faster but less accurate than eye+head, since the eye tracker would decrease the need for head and body rotation.

SUI '17, October 16–17, 2017, Brighton, United Kingdom

ACM ISBN 978-1-4503-5486-8/17/10…$15.00
https://doi.org/10.1145/3131277.3132182

H4: Participants would prefer eye+head over the other two selection techniques since it leverages the advantages of both head- and eye-only selection.

The primary contributions of our work are the first experiment to evaluate eye- and head-based selection performance with the FOVE head-mounted display, and evidence that, contrary to our initial expectations, eye tracking does *not* offer better performance than head-based selection.

2 RELATED WORK

2.1 3D Selection Techniques

Selection techniques include exocentric metaphors and egocentric metaphors. Egocentric metaphors such as virtual hand and ray-based metaphors [23] are in widespread usage today. Techniques like Go-Go [23] "compromise", by combining a virtual hand with arm extension. Image-plane selection [22] is another compromise technique, supporting 2DOF selection of remote objects. Lee et al. [16] compared image-plane selection to a hand-directed ray, head-directed ray, and a ray controlled by both the head and hand, and report that image-plane selection performed best.

Using the eye-tracking capability of the FOVE HMD as a selection technique is perhaps closest to image-plane selection [22]. It requires only 2DOF input, since the user must only fixate on a given pixel. We note that from a technical point of view, this still results in ray-casting, similar to using a mouse for 3D selection. In contrast, head-based selection uses 6DOF movement of the head – although through careful and deliberate head movements, a user could constrain this to just 2DOF rotation. We thus anticipate that eye-tracking could offer superior performance.

Our experiment design is similar to the method proposed by Teather and Stuerzlinger [29, 30] for evaluating 3D selection techniques. Their work extended the ISO 9241-9 standard for use in 3D contexts, using various target depth combinations, and was validated using both a mouse (for consistency with 2D studies using the standard) and various 3D tracking devices.

2.2 Eye Based Interaction

Research on eye based interaction dates to the 1980's [1, 2, 17]. For example, Jacob [13] investigated eye blink and dwell time as selection mechanisms in an effort to overcome the so-called "Midas Touch" problem: subtle eye movements continue to move the cursor, potentially in unintended ways. We avoid this issue by requiring users to press a key on the keyboard to indicate selection.

Starker and Bolt [3] used eye-tracking to monitor and analyze user interest in three-dimensional objects and interface. More recently, Essig et al. [8] implemented a VICON-EyeTracking visualizer, which displayed the 3D eyegaze vector from the eye tracker within the motion-capture system. In a grasping task, their system performed well for larger objects but less so for smaller objects, since a greater number of eye saccades occurred towards boundaries. In testing a variety of objects, they found that a sphere yielded the best results as assessed in a manual annotation task. This was likely because the sphere was bigger than a cup and stapler object, and was not occluded during grasping. These results are consistent with the selection literature, which outlines the importance of target size, distance [26], and occlusion [31].

Lanman et al. [15] conducted experiments using trained monkeys, comparing eye and head movements when tracking moving objects. They report that head movement closely followed the target, while the eye gaze vector was relatively close to the head vector, but moved somewhat erratically. Despite the irregularity of individual eye and head movements, their combination allowed precise target tracking, regardless if the head position was fixed or free. The authors argue that the vestibular system coordinated eye and head motion during tracking, yielding smooth pursuit. These results support our hypothesis that our eye+head selection technique should perform at least as well as head-only selection, while eye-only selection should have the worst accuracy.

Research on eye-only selection conducted by Sibert and Jacob [25] revealed that eye gaze selection was faster than using a mouse. Their algorithm could compensate for quick eye movements, and could potentially be adapted for use in virtual environments. They report that there is also physiological evidence that saccades should be faster than arm movements, which may explain their results. This reinforces our hypothesis that eye-tracking may prove a useful interaction paradigm in VR.

Several performance evaluations of eye-only input have been conducted. Fono and Vertegaal [11] compared four selection techniques, and report that eye tracking with key activation was faster and more preferred than a mouse and keyboard. Vertegaal conducted a Fitts' law evaluation of eye tracking [32] and found that eye tracking with dwell time performed best among four conditions (mouse, stylus, eye tracking with dwell, and eye tracking with click). However, as this study did not employ the standardized methodology for computing throughput (incorporating the so-called accuracy adjustment), the resultant throughput scores cannot be directly compared to other work. Notably, the eye tracker also suffered from a high error rate for both selection methods.

MacKenzie presented an overview of several issues in using eye trackers for input [18]. He also presented the results of two experiments investigating different selection methods using eye tracking, including dwell length, blink, and pressing a key. The eye tracking conditions yielded throughput in the range of 1.16 bits/s to 3.78 bits/s. For reference, ISO-standard compliant studies typically report mouse throughput of around 4.5 bits/s [26]. In these studies, MacKenzie reported mouse throughput of around 4.7 bits/s.

Finally, it is worth noting other applications of eye tracking in immersive VR. Ohshima et al. [21] implemented a gaze detection technique in VEs. Duchowski et al [7] applied binocular eye tracking in virtual aircraft inspection training by recording participants' head pose and eye gaze orientation. Steptoe et al. [28] presented a multi-user VR application displayed in a CAVE. They used mobile eye-trackers to control user avatar gaze direction, with the intent of improving communication between users. They report that participants' gaze targeted the interviewer avatar 66.7% of the time when asked a question. However, eye tracker noise created some confusion as to where participants were looking, contributing to 11.1% of ambiguous cases. We anticipate eye tracker noise may similarly affect our results.

3 METHODOLOGY

3.1 Participants

We recruited eighteen participants (aged 18 to 40, $\mu = 28$ years, 12 male). All participants were daily computer users ($\mu = 5$ hours/day). None had prior experience with eye tracking. Half (nine) had no prior VR experience, five had limited VR experience (having used it once or twice ever), and the rest used VR an average of 5 hours. All participants had colour vision. Fourteen had normal vision, four participants had corrected visions (i.e., they wore corrective lenses). All participants could see stereo imagery, as assessed by pre-test trials.

3.2 Apparatus

Participants wore a FOVE HMD in all trials. See Figure 1. The FOVE display resolution is 2560 x 1440 with a 100° field of view. A unique feature of the display is the two integrated infrared eye-trackers, which offer tracking precision better than 1° at a 120 Hz sampling rate. Like other HMDs, the FOVE offers IMU-based sensing of head orientation, and optical tracking of head position. However, it does not offer IPD correction.

Figure 1. Participant wearing the FOVE HMD while performing the task.

The experiment was conducted on a desktop computer, with an Intel Core i5-4590 CPU, an NVIDIA GeForce GTX 1060 GPU, and 8GB RAM. The experimental interface and testbed was based on discrete-task implementation of the multi-directional tapping test in ISO 9241-9. The software presented a simple virtual environment with spherical targets displayed at the specified depth. See Figure 2. The software was developed using Unity 5.5 and C#.

3.3 Procedure

The experiment took approximately 40 minutes in total for each participant. Participants were first briefed on the purpose and objectives of the experiment, then provided informed consent before continuing.

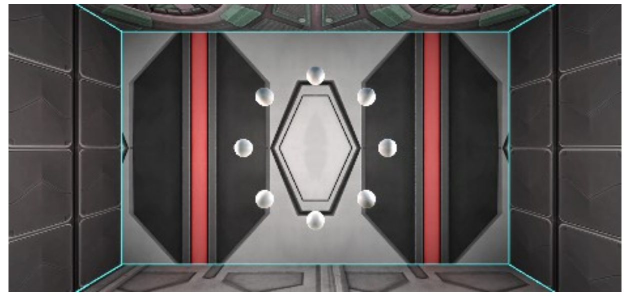

Figure 2. Software used in the experiment depicting the selection task.

Upon starting the experiment, participants sat approximately 60 cm from the FOVE position tracker, which was mounted on the monitor as seen in Figure 1. They first completed the FOVE calibration process, which took approximately one minute. Calibration involved gazing at targets that appeared at varying points on the display. This calibration process was also used as pre-screening for the participants: Prospective participants who were unable to complete the calibration process were disqualified from taking part in the experiment. Prior to each new condition using the eye tracker (i.e., eye-only and eye+head), the eye tracker was re-calibrated to ensure accuracy throughout the experiment. Following calibration, the actual experiment began.

The software presented eight gray spheres in circular arrangement in the screen centre. See Figure 2. Participants were instructed to select the orange highlighted sphere as quickly and accurately as possible. Selection involved moving the cursor (controlled by either the eye tracker or head orientation) to the orange sphere and pressing the "z" key. The participant's finger was positioned on the "z" key from calibration to the experiment's end to avoid homing/search for the key. Alternative selection indication methods would also influence results (e.g., fixating the eye on a target for a specified timeout would decrease selection speed and thus also influence throughput [19]). We note that Brown et al. [6] found no significant difference between pressing a key and a "proper" mouse button in selection tasks. However, our future work will focus on alternative selection indication methods.

Upon completing a selection trial, regardless if the target was hit or missed, the next target sphere would highlight orange. A miss was determined by whether the cursor was over the target or not when selection took place. Software logged selection coordinates, whether the target was hit, and selection time. Upon completion of all trials, participants completed a 7-point questionnaire based on ISO 9241-9 and were debriefed in a short interview.

3.4 Design

The experiment employed a 3 × 3 × 4 within-subjects design. The independent variables and their levels were as follows:

Input Method: Eye-only, head-only, eye+head
Target Width: 0.25m, 0.5 m, 0.75 m
Target Depth: 5 m, 7 m, 9 m, mixed

With eye-only, the FOVE head tracker was disabled. With head-only, the FOVE eye tracker was disabled and the cursor was fixed in the screen centre. The eye+head input method used both the eye and head trackers, and represents the "default" usage of the FOVE. Although eye-only does not represent typical usage of the FOVE, it was included to provide a reasonable comparison point to previous eye-tracking Fitts' law studies [18].

Three target sizes yielded three distinct indices of difficulty, calculated according to Equation (1). We used three fixed depths, plus mixed depths to add a depth component to the task. In the fixed depth conditions, all targets were presented at the same depth (5, 7, or 9 m from the viewer). In the mixed depth conditions, the sphere at the 12 o'clock position (the top sphere) was positioned at a depth of 5 m. Each subsequent sphere in the circle (going clockwise) was 10 cm deeper than the last. See Figure 3.

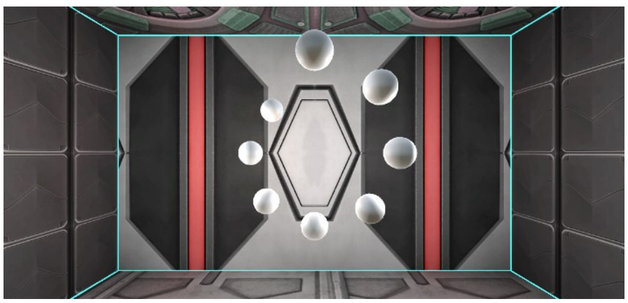

Figure 3. Same-sized spheres in a mixed depth configuration.

All three target widths were crossed with all four depths, including mixed depth. The ordering of input method was counterbalanced according to a Latin square. There were 15 selection trials per combination of target depth and target width, hence the total number of trials was 18 participants × 3 input methods × 4 depths × 3 widths × 15 trials = 9720 trials.

The dependent variables included throughput (in bits/s, calculated according to Equation (2)), movement time (ms), and error rate (%). Movement time was calculated as the time from the beginning of a trial, to the time the participant pressed the "z" key, which ended the selection trial. Error rate was calculated as the percentage of trials where the participant missed the target.

Based on previous work [14, 27], we calculated *ID* using rotation angle between targets for distance, and the angular size of the target. *ID* was calculated as follows:

$$ID = \log_2\left(\frac{\alpha}{\omega} + 1\right) \quad (1)$$

where α is the rotation angle from sphere B to sphere A and ω is the angular size the target sphere (i.e., angular interpretations of *A* and *W*). Then, throughput was calculated as:

$$TP = \frac{ID}{MT} \quad (2)$$

where *MT* is the average movement time for a given condition. Angular measures for distance and target size (α and ω) were derived trigonometrically, see Figure 4.

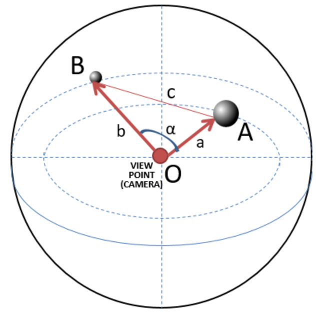

Figure 4. The same-sized spheres A and B at different depths form triangle ∠AOB with the view point O. Although the straightline distance between A and B is C, the angular distance is represented by α. A similar calculation is used for the angular size of targets from the viewpoint.

Finally, we also collected subjective data via nine questions using a 7-Likert scale. These questions were based on those recommended by ISO 9241-9.

4 RESULTS AND ANALYSIS

Results were analyzed with repeated measures ANOVA.

4.1 Error Rates

Mean error rates are summarized in Figure 5. There was a significant main effect of input method on error rate ($F_{2,14} = 13.99$, $p < .05$). The Scheffé post-hoc test revealed that the difference between all three input methods was significant ($p < .05$). Eye-only and eye+head had much higher errors than head at roughly 40% and 30% vs. 8% respectively. The high standard deviation reveals great variation in performance, especially for eye-only and eye+head. This suggests that participants had much greater difficulty selecting targets with the eye tracker, consistent with previous results [32].

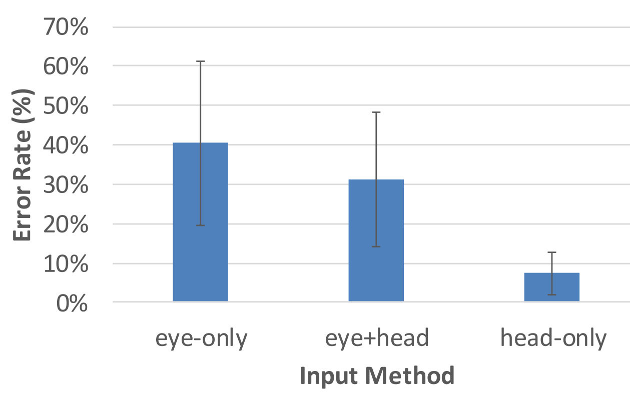

Figure 5. Mean error rates for each input method. Error bars show ±1 *SD*.

Figure 6 depicts error rates by target depth and size for each input method. Note that error rate increased for both smaller targets, and targets farther from the viewpoint. The error rates of eye-only and eye+head increased sharply, while error rates of head-only increased only slightly. Eye-only and eye+head had a varied greatly depending on the target size and depth. Eye-only and eye+head were notably worse with the deepest target depth (9 m). The effect of target size – expected in accordance with Fitts' law – was also quite pronounced with mixed-depth targets.

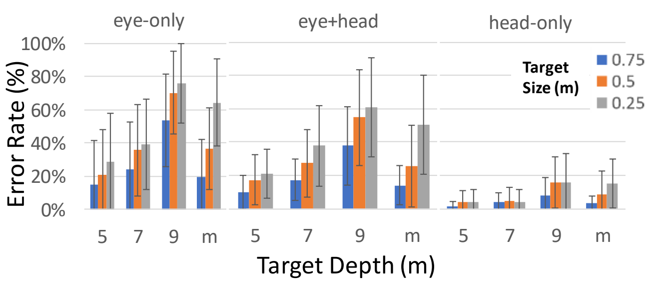

Figure 6. Error rate by target size and target depth for each input method. Note: 'm' depth represents mixed depths. Error bars show ±1 *SD*.

We note that the angular size of the target combines both target depth and size, both factors which influence error rates, as seen above. Due to perspective, a farther target will yield a smaller angular size, and according to Fitts' law, should be a more difficult target to select [30]. Hence, we also analyzed error rates by angular size of the targets. As expected, angular size had a dramatic effect on selection accuracy. As seen in Figure 7, we detected a threshold of about 3°. Targets smaller than this (either due to presented size, depth, or their combination) are considerably more difficult to select with all input methods studied – but especially for the eye-only and eye+head input methods. We thus suggest ensuring that selection targets are at least 3° in size, to maximize accuracy.

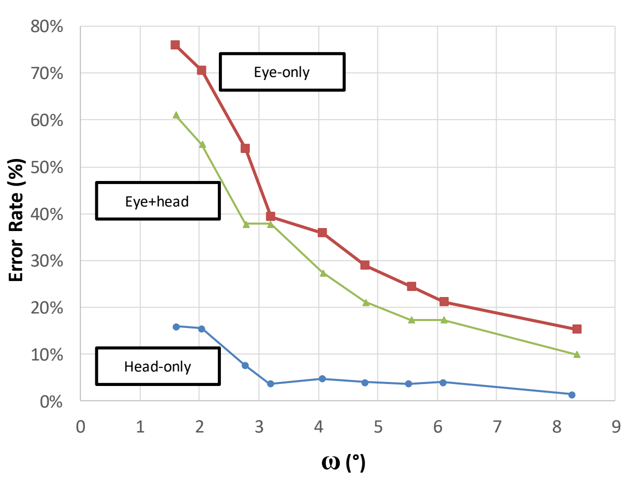

Figure 7. Average error rate for each input method vs. angular size of the target (ω), in degrees.

4.2 Movement Time

Mean movement times are summarized in Figure 8. There was a significant main effect of input method on the movement time ($F_{2,14} = 4.71$, $p < .05$). The Scheffé post-hoc test revealed significant differences between head-only and the other two input methods ($p < .05$). Eye+head and eye-only were not significantly different from each other. This again suggests that the presence of eye tracking yielded worse performance – the one input method that did not use it (head-only) was significantly faster than both input methods that did.

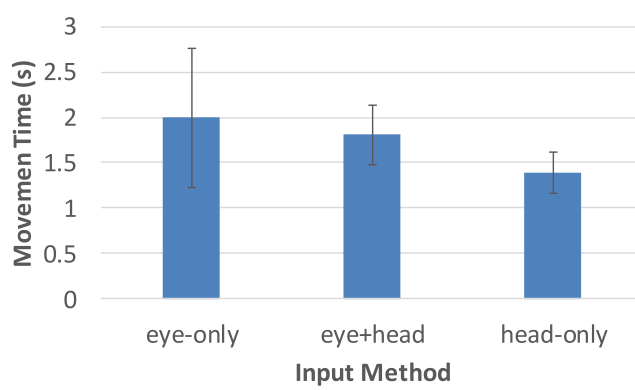

Figure 8. Movement time by selection method. Error bars show ±1 *SD*.

As seen in Figure 9, movement time increased slightly as the target size became smaller. However, the effect of target depth was more pronounced, particularly with the eye-only input method. The other two input methods increased slightly and similarly.

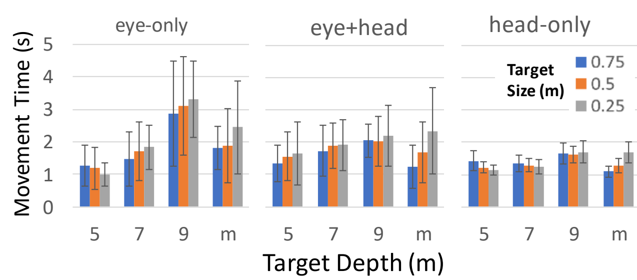

Figure 9. Movement time by target size and depth for each selection method. Note: 'm' depth represents mixed depths. Error bars show ±1 *SD*.

4.3 Throughput and Fitts' Law Analysis

Throughput scores are summarized in Figure 10. There was a significant main effect for input condition on throughput ($F_{2,14} = 21.99$, $p < .05$). The Scheffé post-hoc test also showed significant differences ($p < .05$) between eye+head and head-only, and head-only and eye-only. However, eye+head and eye-only were not significantly different, which again suggested some difference due to the presence of eye tracking. Head-only was once again the best among the three input methods. The throughput scores of eye-only and eye+head were in the range reported by Mackenzie [18], yet notably lower than average throughput for the mouse [26]. We note that throughput was also somewhat higher than that reported by Teather and Stuerzlinger [29, 30] for a handheld ray-based selection technique.

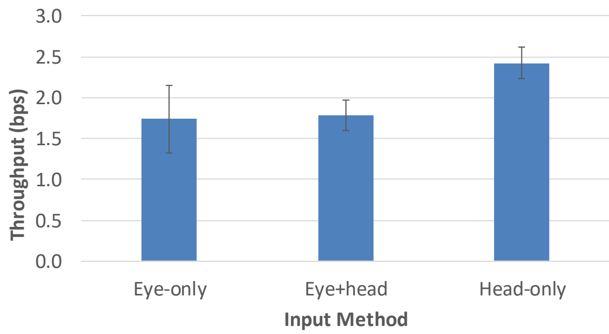

Figure 10. Throughput by input methods.

As is common practice in Fitts' law experiments, we produced linear regression models for each selection method showing the relationship between *ID* and *MT*. These are shown in Figure 11.

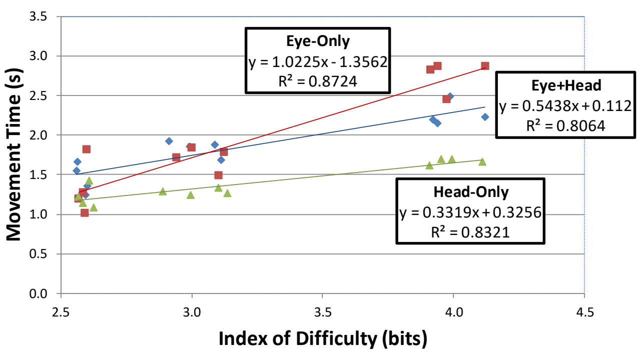

Figure 11. Regression models for all input methods.

Note that the presented R^2 scores are quite high, ranging from between 0.8 and 0.87. This suggests a fairly strong predictive relationship between *ID* and *MT*, which is typical of interaction techniques that conform to Fitts' law. We note that these scores are somewhat lower than in other research using input devices like the mouse [26], but in line with previous research on 3D selection [29, 30]. Interestingly, the eye-only input method offered the best fitting model, suggesting that eye-tracking conforms to Fitts' law better than head-based selection [18, 32].

4.4 Subjective Questionnaire

The device assessment questionnaire consisted of 9 items, modelled after those suggested by ISO 9241-9. We asked each question for each input method. Each response was rated on a 7-point scale, with 7 as the most favourable response and 1 the least favourable response. Responses are seen in Figure 12.

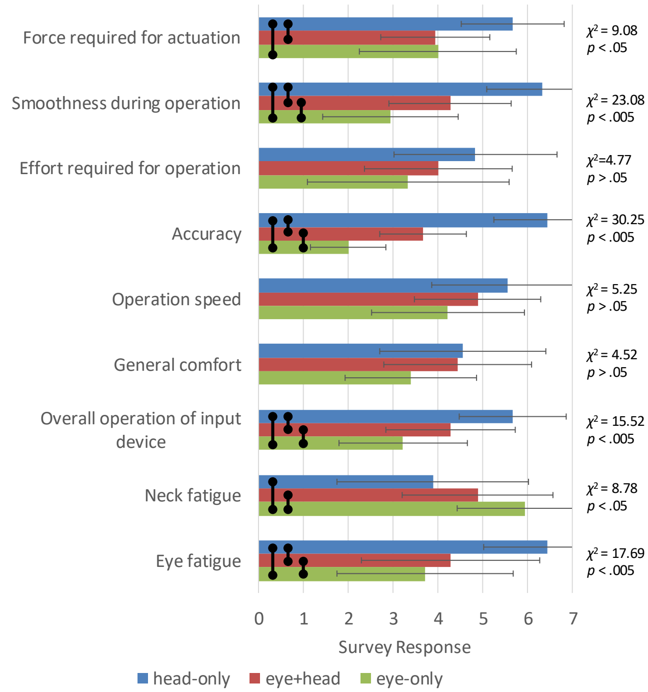

Figure 12. Average of response scores for each survey question. Error bars show ±1 SD. Higher scores are more favourable in all cases. Statistical results via the Friedman test shown to the right. Vertical bars (●—●) show pairwise significant differences per Conover's F test posthoc at the p <0.05 level.

Overall, participants rated head-only best on all points expect neck fatigue. Eye-only was rated best on neck fatigue. Conversely, and perhaps unsurprisingly, head-only was rated best on eye fatigue, and eye-only was rated worst. Participants were also aware of the accuracy difference between the input methods; they reported head-only was most accurate, followed by eye+head, with eye-only rated worst, much like the error rate results shown earlier.

4.5 Interview

Following completion of the experiment, we debriefed the participants in a brief interview to solicit their qualitative assessment of the input methods. Eleven participants preferred head-only because it provided high accuracy, and it was the most responsive and comfortable. Six participants found eye-only the worst, reporting that it was difficult to use. Some indicated that due to their prior experience wearing standard HMDs, they were already used to head-based interaction, which may help explain their preference towards head-only. However, they tended to indicate that they found eye-only inefficient.

Five participants found eye+head the worst. Much like our initial hypothesis, at the onset of the experiment, these participants expected eye+head would offer better performance, but were surprised to find that it did not. A few participants indicated that they experienced some nausea and neck fatigue with eye+head. Finally, five participants rated eye-only the best. Although it did not provide accurate operation, these participants felt comfortable using it. They also complained about neck fatigue with both head-based input methods, and indicated that they looked forward to wider availability of eye-tracking HMDs in the future. Some even suggested that for tasks that did not require precision, they would always choose eye-tracking.

5 Discussion and Future Work

Before the experiment, we hypothesized that using eye and head tracking together – the eye+head input method – would offer the best performance of the three input methods, since it offered the best capabilities of both eye- and head-tracking. Our data, however, disproved this hypothesis. In fact, the head-only input method performed the best across all dependent variables, especially accuracy. In contrast, the two input methods utilizing eye tracking (eye-only and eye+head) were fairly close in performance, with eye+head generally performing better than eye-only. We hypothesized that head-only would yield the lowest error rates; this hypothesis was confirmed. We also hypothesized that participants would prefer eye+head, but this was not the case. Based on past work, we had expected that eye+head would provide a selection method consistent with how we use our eyes and head together in pursuit tracking [15]. However, during testing, we observed that the cursor sometimes jittered, resulting in poor precision with eye+head. This may be a limitation of the hardware.

Previous eye tracking research relates the importance of calibration problems, which can drastically influence the data [1, 2, 13]. Two potential participants were excluded because despite 5 attempts, they still failed the calibration. This might be an inherent flaw of FOVE's calibration algorithm or hardware. We also observed that calibration quality greatly influenced selection performance. For example, during the calibration phase, participants had to follow a moving green dot with their eye gaze. One participant mentioned that the green dot stopped moving for more than 3 seconds on the 9 o'clock and 1 o'clock direction. This may be due to a software bug, or because the eye tracker had difficulty detecting the participant's eyes. As a result, during testing, that participant could not reliably select targets in those directions, necessitating re-calibration of the eye tracker. In all these sessions, although the participant had passed the calibration component, such pauses during the calibration process could still yield poor results, likely affecting performance with both eye-tracking input methods. Participants suggested improving the calibration process in future, which may yield better results with the eye-only and eye+head input methods.

As detailed above, participants strongly favoured the head-only input method. In the eye-only and eye+head sessions, participants indicated that they could comfortably and reliably select larger spheres. However, when spheres were smaller and/or deeper into the scene (i.e., smaller in terms of angular size), participants felt very frustrated and uncomfortable, particularly when missing the targets. Based on this observation, and our earlier analysis of angular sizes, we recommend designers to avoid targets smaller than 3° in size. While it is well-known that target size influences pointing difficulty [9, 26], this seems especially important with eye-only selection. In contrast, large targets and relatively closer targets are considerably easier for participants to select.

Interestingly, during the interview most (16/18) participants felt that eye+head would work well in VR first-person shooter games, despite the largely negative results yielded by this input method. Participants suggested that head-only could cause sickness, and eye-only is too inaccurate. Participants suggested that an improved version of eye+head would work well for shooting. Similarly, half felt eye-only would work well for menu selection, while the rest thought head-only would work best. One suggested that assuming large enough widgets, any technique would be effective.

6 CONCLUSIONS

It seems likely that eye-tracking will become available in more head-mounted displays in the near future. While eye tracking has been used previously to support selection tasks [11, 18, 25, 32], our study is the first to look at eye-only selection performance in VR environment using Fitt's law and ISO 9241-9. We found that head-only selection offered the fastest selection times and the best accuracy. Moreover, it was strongly preferred by participants. The combination of eye-tracking and head-based selection (our eye+head input method) performed roughly between the other two, failing to leverage the benefits of each. Our results indicate that, at least for the time being and in the absence of more precise eye trackers with better calibration methods, head-only selection is likely to continue to dominate VR interaction.

A limitation of this study is that we necessarily constrained our test conditions to conform to the Fitts' law paradigm. This included keeping participants seated – although we note that seated VR is still a major use case, e.g., gaming on the Oculus *Rift*. We also constrained targets to only appear in front of the viewer, which is somewhat unrealistic. We considered having targets outside the field of view, but this would not be a Fitts' law task as it would incorporate a search task as well as selection. Future work will focus on eye-based interaction in VR using a broader range of tasks (e.g., navigation, manipulation) and enhanced task realism (e.g., selecting targets outside the field of view).

ACKNOWLEDGEMENTS

Thanks to all participants. This work was supported by NSERC.

REFERENCES

[1] Bolt, R. A. 1981. Gaze-orchestrated dynamic windows. In *ACM SIGGRAPH Computer Graphics* (Vol. 15, No. 3), 109-119.

[2] Bolt, R. A. 1982. Eyes at the interface. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '82). ACM, New York, 360-362.

[3] Bolt, R. A. 1990. A gaze-responsive self-disclosing display. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '90), ACM, New York, 3-10.

[4] Bowman, D. A., & Hodges, L. F. 1997. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments.

In *Proceedings of the ACM Symposium on Interactive 3D Graphics* (SI3D), ACM, New York, 35-38.
[5] Bowman, D. A., Kruijff, E., LaViola Jr, J. J., & Poupyrev, I. 2001. An introduction to 3-D user interface design. *Presence: Teleoperators and Virtual Environments*, *10*(1), 96-108.
[6] Brown, M. A., & Stuerzlinger, W. 2014. The performance of un-instrumented in-air pointing. In *Proceedings of Graphics Interface 2014*, Canadian Information Processing Society, Toronto, 59-66.
[7] Duchowski, A. T., Medlin, E., Gramopadhye, A., Melloy, B., & Nair, S. 2001. Binocular eye tracking in VR for visual inspection training. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (VRST '11). ACM, New York, 1-8.
[8] Essig, K., Dornbusch, D., Prinzhorn, D., Ritter, H., Maycock, J., & Schack, T. 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). ACM, New York, 37-44.
[9] Fitts, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381-391.
[10] Foley, J. D., Wallace, V. L., & Chan, P. 1984. The human factors of computer graphics interaction techniques. *IEEE Computer Graphics and Applications*, *4*(11), 13-48.
[11] Fono, D., & Vertegaal, R. 2005. EyeWindows: Evaluation of eye-controlled zooming windows for focus selection. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '05). ACM, New York, 151-160.
[12] ISO, Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices (ISO 9241-9), International Organisation for Standardisation. Report Number ISO/TC 159/SC4/WG3 N147, February 15, 2000.
[13] Jacob, R. J. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '90). ACM, New York, 11-18.
[14] Kopper, R., Bowman, D. A., Silva, M. G., & McMahan, R. P. 2010. A human motor behavior model for distal pointing tasks. *International Journal of Human-Computer Studies*, *68*(10), 603-615.
[15] Lanman, J., Bizzi, E., & Allum, J. 1978. The coordination of eye and head movement during smooth pursuit. *Brain Research*, *153*(1), 39-53.
[16] Lee, S., Seo, J., Kim, G. J., & Park, C. M. 2003. Evaluation of pointing techniques for ray casting selection in virtual environments. In *the International Conference on Virtual Reality and its Application in Industry* (Vol. 4756, No. 1), 38-44.
[17] Levine, J. L. 1981. *An eye-controlled computer*. IBM Research Division, TJ Watson Research Center.
[18] MacKenzie, I. S. 2011. Evaluating eye tracking systems for computer input. In Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies, 205-225.
[19] MacKenzie, I. S., & Teather, R. J. (2012). FittsTilt: the application of Fitts' law to tilt-based interaction. In *Proceedings of the ACM Nordic Conference on Human-Computer Interaction* (NORDICHI '12). ACM, New York, 568-577.
[20] Mine, M. R. (1995). Virtual environment interaction techniques. *UNC Chapel Hill CS Dept*.
[21] Ohshima, T., Yamamoto, H., & Tamura, H. 1996. Gaze-directed adaptive rendering for interacting with virtual space. In *Proceedings of the IEEE Virtual Reality Annual International Symposium.* IEEE, New York, 103-110.
[22] Pierce, J. S., Forsberg, A. S., Conway, M. J., Hong, S., Zeleznik, R. C., & Mine, M. R. 1997. Image plane interaction techniques in 3D immersive environments. In *Proceedings of the Symposium on Interactive 3D Graphics* (SI3D). ACM, New York, 39-43.
[23] Poupyrev, I., Billinghurst, M., Weghorst, S., & Ichikawa, T. 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (UIST '96). ACM, New York, 79-80.
[24] Poupyrev, I., Weghorst, S., Billinghurst, M., & Ichikawa, T. 1997. A framework and testbed for studying manipulation techniques for immersive VR. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (VRST '97). ACM, New York, 21-28.
[25] Sibert, L. E., & Jacob, R. J. 2000. Evaluation of eye gaze interaction. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI 2000). ACM, New York, 281-288.
[26] Soukoreff, R. W., & MacKenzie, I. S. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, *61*(6), 751-789.
[27] Stoelen, M. F., & Akin, D. L. 2010. Assessment of Fitts' law for quantifying combined rotational and translational movements. *Human Factors*, *52*(1), 63-77.
[28] Steptoe, W., Wolff, R., Murgia, A., Guimaraes, E., Rae, J., Sharkey, P., Roberts, D., and Steed, A. 2008. Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '08). ACM, New York, 197-200.
[29] Teather, R. J., & Stuerzlinger, W. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. In *Proceedings of the IEEE Symposium on 3D User Interfaces* (3DUI '11)*,* IEEE, New York, 87-94.
[30] Teather, R. J., & Stuerzlinger, W. 2013. Pointing at 3D target projections with one-eyed and stereo cursors. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, 159-168.
[31] Vanacken, L., Grossman, T., & Coninx, K. 2007. Exploring the effects of environment density and target visibility on object selection in 3D virtual environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces* (3DUI '07). IEEE, New York, 117-124.
[32] Vertegaal, R. 2008. A Fitts law comparison of eye tracking and manual input in the selection of visual targets. In *Proceedings of the International Conference on Multimodal Interfaces.* ACM, New York, 241-248.
[33] Zhai, S., & Milgram, P. 1994. Input techniques for HCI in 3D environments. In *Conference Companion on Human Factors in Computing Systems*. ACM, New York, 85-86.
[34] Zhang, X., & MacKenzie, I. 2007. Evaluating eye tracking with ISO 9241-part 9. *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, 779-788.