Low-Fi VR Controller: Improved Mobile Virtual Reality Interaction via Camera-Based Tracking

Kristen Grinyer
Carleton University, Ottawa, Canada
kristengrinyer@cmail.carleton.ca

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

Figure 1: Left: Low-Fi VR Controller in right hand and cardboard piece with selection markers in left hand depicting the Marker selection method. Right: VR view of the controller with sphere and cube objects indicating tracking statuses of markers

ABSTRACT

Mobile virtual reality (VR) provides an accessible alternative to high-end VR, but currently offers only limited interaction, hindering its usability, the variety of VR experiences it can provide, and widespread adoption. To improve interaction on mobile VR, we present a novel input solution: the Low-Fi VR Controller. The controller uses the smartphone camera to track markers to provide 6DOF input. It costs virtually nothing as it is made of cost-effective accessible materials. We performed a user study based on Fitts' law to evaluate the controller's performance in selection tasks, and to compare three selection activation methods (Instant, Dwell, Marker). Despite tracking issues, selection throughput with the Instant method was comparable to other similar ray-based selection techniques reported in other studies, at roughly 2.2 bps. Our results validate the controller as an acceptable 3D input device and will propose avenues to improve performance and user experience with the controller in future work.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **Interaction devices**;

KEYWORDS

Fitts' law, Google Cardboard, Pointing, Selection, Input device

1 INTRODUCTION

Virtual reality (VR) offers improved, immersive experiences in education [3, 61] training [58], mental and physical health interventions [50], various therapies [59], collaborative work [65], and entertainment [52]. Noting this, many major tech companies (e.g., Meta, Apple) have begun investing in VR and manufacturing commercial technologies meant for everyday use with a focus on entertainment, work, and social experiences. The technology is poised to become ubiquitous in the next decade. Yet, VR technology is not universally available. Modern VR displays with tracked controllers range from $300-$1,400 USD. These devices are prohibitively expensive for many users [12], leaving behind segments of the population who could benefit from the technology [57].

A potential low-cost alternative, mobile VR (MVR), emerged even before the current generation head-mounted displays, with early projects such as FOV2GO [41] proposing mobile phones as a VR platform. Exemplified by Google's Cardboard, MVR has provided a low-cost solution for the broader public to experience VR [1, 30, 62]. Typically, MVR headsets range from $10-20 USD and are a head-mounted display (HMD) "shell"; the user inserts their smartphone, which acts as a computing device and provides head rotation via the phone's internal sensors. Between the low cost of these devices and the ubiquity of smartphones [42, 48], MVR has the potential to make VR technology more widely available than high-end VR devices.

While mobile VR has alleviated cost as a barrier to VR use, MVR devices lack the tracking capabilities offered by higher-fidelity devices (e.g., Meta Quest). They do not include tracked controllers, hand-tracking, or room-scale tracking, which are largely required

to support effective 3D interaction in VR. Bluetooth controllers can be purchased for $25-30 USD, but since only their rotational movements are tracked they cannot provide as complex interactions as high-end controllers. While some MVR experiences use a gaze-fixed cursor to acquire targets, most MVR applications are limited to passive entertainment [45] where users can look around the virtual space [30], but generally cannot interact or actively explore the virtual environment (VE) [45]. As it stands, MVR cannot support applications with proven benefits to health, social, and professional activities [3, 50, 52, 58, 59, 61, 65].

Noting MVR's potential benefits despite this "interaction gap", we propose a novel MVR input device, the Low-Fi VR Controller, to improve the input capabilities of low-fi devices at virtually no extra cost to users. Our proposed solution (see Figure 1) employs the camera universally available on all modern smartphones to track an external controller made of multiple paper markers attached to a piece of cardboard. It provides 6 degree of freedom (DOF) tracking capabilities: a standard in high-end VR controllers. Currently, selection is performed via raycasting, but this could be expanded to other selection techniques (e.g., virtual hand) as well.

To evaluate the efficacy of our controller, we conducted a user study comparing the user performance and user experience of the controller. An outstanding issue with our controller design is how to indicate selection, i.e., "click" the button; we thus tested the device with three different selection activation methods: Instant, Dwell, and Marker. Each is fully described along with full study details in section 4. The analysis and discussion primarily focused on comparing these three methods. Participants performed selection tasks conforming to the ISO-9241-9 standard evaluation methodology. To further evaluate the controller, participants filled out a questionnaire on input device comfort and provided additional subjective data.

Our goals were to determine whether our Low-Fi VR Controller offers basic selection performance and get user perspectives to identify potential improvements in future iterations. An eventual goal of our work is to further improve the controller quality and compare its performance to a high-end VR controller and the standard input technique of MVR: head-gaze-based selection [45]. Our research questions included:

R1: What is the performance potential of the Low-Fi VR Controller in terms of selection time and throughput?

R2: Which selection activation method offers the best performance in terms of selection time and throughput?

R3: What improvements can be made to the design of the controller to improve its performance and usability?

The main contributions of our work include:

- The design of our Low-Fi VR Controller, which provides 6DOF input for interaction in mobile/cardboard VR via low-cost camera-based marker tracking
- Further insight and design considerations for selection activation methods with the controller
- A formal experiment providing evidence that the controller offers reasonable performance in selection tasks

2 RELATED WORK

2.1 Improving Interaction in Mobile VR

The main challenge of MVR is its limited interaction capabilities, which largely prevents MVR from accessing the kinds of experiences and benefits offered by high-fidelity VR devices [1, 16, 22, 25, 29, 30, 45]. We focus on improving selection, so we review selection interactions in MVR. Selection in MVR is typically performed using head-gaze-based selection with a reticle pointer centered in the view. Selection indication occurs either by dwell or a lever on the cardboard HMD [25]. There have been many proposed input techniques in an effort to improve interactions in MVR [13, 16, 20, 25, 29, 44, 56, 60]. However, proposed solutions either offer only very limited performance compared to high-fidelity VR or require an additional expensive device contradictory to the low-cost accessible nature of MVR. We describe various approaches below.

2.1.1 Additional Hardware Solutions. Several researchers have investigated the use of smartwatches for ray-based selection in MVR [20, 25]. The smartwatch inertial sensors control the ray orientation, which originates at the smartphone camera position. While the smartwatch provides reliable rotational tracking, these solutions are limited to 3DOF control and a fixed origin ray. This solution assumes the user already owns a smartwatch, which runs counter to the cost-accessible quality of MVR.

Other researchers have proposed using a second smartphone to create a 6DOF tracked controller [27, 36]. The camera on the HMD is used to track the second smartphone's position, while the second smartphone's inertial sensors provide orientation. Mohr et al. [36] further used the camera on the handheld smartphone to track a marker placed in front of the HMD for additional positional tracking information. This allowed their controller to be usable outside of the HMD camera's field of view (FOV). While these solutions provide comparable movement freedom to a high-end VR controller, the need for a second smartphone makes them impractical for consumer use.

Others have attached hardware to the front surface of the cardboard HMD for input including a touch screen with a one-to-one mapping to the FOV [16], and physical buttons that were dynamically placed in the same location as virtual buttons for direct manipulation [56]. While effective, these approaches again require extensive additional hardware beyond the cardboard viewer and single smartphone.

2.1.2 Low-Cost Solutions. Novel input techniques have been proposed with cost in mind but the input they provide is still noticeably limited. ScratchVR [29] used an additional cardboard piece with a magnetic washer on the side of the HMD that enabled bidirectional scrolling in MVR. While this improved menu interactions, it is not an ideal solution for general purpose target selection.

Other studies have focused on gestures to improve MVR interaction. Temporal signals made from tap gestures on the cardboard HMD surface can be recognized by the smartphone's motion sensors [60]. Similarly, the acoustic signals of sliding gestures on the HMDsurface can be detected by the smartphone's microphone [13].

Previous work has explored camera-based approaches to interaction in mobile VR [10, 30, 63]. Luo et al. [30] compared ray-based

and direct touch selection techniques with different selection activation methods and found camera-based tracking in MVR is best suited for ray-based interaction with a button for selection indication. The present study assesses dwell and marker occlusion as potential selection activation methods for MVR. Other studies have used external servers to offload expensive tasks from the smartphone [10, 63] using the smartphone's camera to capture data to send to a nearby server. The server then sends back pose information used for hand or body tracking. This approach improves the quality of MVR experiences while only needing an additional PC for operation. To our knowledge, there has been no proposed low-cost solution that provides 6DOF pointing interaction similar to high-fidelity VR. Our work aims to bridge this gap in the current literature.

2.2 Selection in Virtual Reality

After locomotion, selection is among the most common fundamental tasks in VR and a common precursor for other fundamental tasks (e.g., manipulation, system control) [2]. A well-designed selection technique is imperative to a positive user experience, thus we investigate the use of the Low-Fi VR Controller in selection tasks. Among the most commonly employed selection techniques in VR is remote pointing via raycasting. It is favoured over the similarly common virtual hand metaphor [2] as it allows users to select objects outside their reach and requires less physical movement to operate [2, 28]. Multiple studies have found remote pointing offers better selection performance than other selection metaphors [2].

In general, humans have difficulty operating in 3D VEs as they lack an understanding of 3D spatial relationships and operating multiple DOFs at once [2]. Unlike operating a mouse in 2D, handheld controllers for 3D interaction require complex arm movements using larger, slower muscles [9, 26]. Previous work compared the selection performance between a mouse, 6DOF controller, and optically tracked 3D pen in VR using the ISO-9 methodology [43]. The pen offered better objective performance than the controller and was preferred by participants. This is likely because the participants were more familiar with holding a pen-like device, and they used smaller, faster muscles to control the pen [5, 64].

Most selection techniques require some kind of "click" or equivalent event to indicate selection [7]. Mutasim et al. [40] compared click, 300 ms dwell, and pinch gesture activation using eye-gaze-based pointing in a selection task. On average, dwell had the slowest selection time and lowest throughput score, but yielded significantly fewer errors than click or pinch. Click was fastest and most preferred by users. Lowering the dwell time (potentially to 0, i.e., instant selection) to increase selection speed introduces the Midas Touch problem [24] of unintended selections; such techniques are generally impractical for real-world usage [34].

2.3 Fitts' Law

We employ the ISO 9241-9 standard for our evaluation, which is based on Fitts' law. Fitts' law is a predictive model of human performance when performing rapid aimed movements such as selection tasks [32]. Given target width (W) and amplitude (A), the distance to the target, Fitts' law models the relationship between movement time (MT) and index of difficulty (ID) [4]. ID is measured in bits,
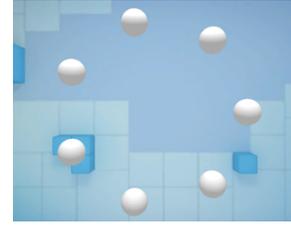
Figure 2: ISO-9241-9 selection task. Spherical targets in a ring formation evenly spaced apart.

and a and b are the intercept and slope of the linear regression line illustrating this relationship [15].

$$MT = a + b\,ID \tag{1}$$

$$ID = \log_2\left(\frac{A}{W} + 1\right) \tag{2}$$

Fitts' law is also used to quantify human performance in terms of throughput (TP). Throughput combines speed and accuracy to measure performance in a target acquisition task and has been shown to be consistent despite speed/accuracy tradeoffs common to such tasks [33]. This property, along with its remarkable consistency between studies make it a valuable and preferred option for comparing selection techniques versus simply using movement time or accuracy [53]. It is independent of A and W, since as ID changes, MT changes inversely [32]. Throughput is defined as:

$$TP = ID_e/MT \tag{3}$$

where ID_e is the effective index of difficulty (equation 4). ID_e adjusts W based on selection distribution to yield effective width (W_e). W_e adjusts the presented width to 4.133 standard deviations of selection coordinates around the target centre; this corresponds to fixing the experimental error rate to 4%. The effective amplitude, A_e (equation 4), is the average of the actual distance the cursor moves for each selection. Both effective measures better represent the task participants perform, and adjust the experiment accuracy to 4%, thus decreasing susceptibility of throughput to speed/accuracy tradeoffs [33, 53].

$$ID_e = \log_2\left(\frac{A_e}{W_e} + 1\right) \tag{4}$$

The ISO 9241-9 standard [23] prescribes the use of throughput as a primary dependent variable, and is typically employed in 2D scenarios. We employ this standard task (see Figure 2) in our experiment, and note that we calculate and report throughput according to two variations. One projects the closest point on the selection ray onto the task axis (line between subsequent targets), treating this as the cursor, effectively providing a 1D SD_x value in accordance with the ISO standard. The other, proposed by Teather and Stuerzlinger [54] uses the straight-line 3D distance from the target centre to the selection coordinate, which penalizes inaccuracy in depth. There have been many proposed 3D extensions to Fitts' Law over recent decades [6, 11, 14, 17, 31, 38, 49]; calculating throughput for 3D tasks remains an ongoing topic of discussion and is beyond the scope of this paper.
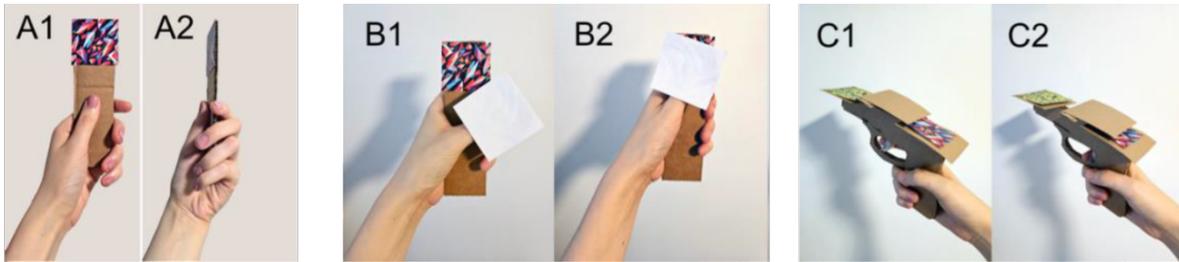
Figure 3: A1, B1, and 1C show the Flip, Thumb-over, and Trigger marker occlusion techniques respectively. A2, B2, and C2 demonstrate how a selection is activated in each technique.

3 DESIGN OF THE LOW-FI VR CONTROLLER

Prior to evaluating the Low-Fi VR Controller, we devised several prototypes and informally tested them to provide a general proof of concept. This included testing various software and physical setups to find the optimal design for the controller prototype to offer 6DOF tracking via low-cost materials.

3.1 Choosing the Software

Previous work that tracked a secondary smartphone via a marker displayed on its screen [13] used the open-source framework So-lAR[1]. SolAR provides camera-based marker tracking in Unity in conjunction with the 2019 Google VR (GVR) SDK. However, So-lAR is only capable of tracking one marker at a time. To enhance tracking quality, we instead explored other AR frameworks but had issues with combining these AR frameworks with a stereo-scopic view (provided either by the GVR SDK or a custom one); consistently, the stereo view and the AR framework being tested had conflicting project setting requirements. After testing many software version combinations, we were able to integrate Vufo-ria v.10.15.3 with the open framework Google XR Plugin v.1.2 in Unity 2021. We found Vuforia provided more reliable image track-ing than other frameworks (like ARCore) and it comes with more relevant features to this project. Specifically, Vuforia's Multi Target feature[2] allowing developers to use physical 3D markers comprised of multiple 2D images.

3.2 Pilot Study

We tested and compared multiple marker configurations and se-lection activation methods when designing our controller proto-type. Initially, we used a single tracked image marker attached to a piece of cardboard representing the controller. Selection activa-tion used a 300 ms dwell on the target. We note that a discrete selection activation event (e.g., a button press) is likely preferable to dwell, as it gives the user more freedom over what, when, and where an object is selected. We thus implemented a second se-lection activation method using a second tracked marker. While this second marker was in-view of the camera, the user could pre-pare their selection position by pointing the selection ray at the desired target. Occluding the second marker (e.g., covering it by the hand, rotating it out of view), would activate selection. We refer to

---

[1]https://solarframework.github.io
[2]https://library.vuforia.com/objects/multi-targets

this second marker as the selection marker. We conducted a pilot study comparing different marker occlusion techniques to deter-mine the best approach to hiding the selection marker from the camera.

The pilot study compared three marker occlusion techniques: Flip, Thumb-over, and Trigger. With Flip, the selection marker was attached to a piece of cardboard. To indicate selection, the user simply 'flipped' the cardboard piece away from camera view by turning the cardboard 90 degrees left or right (see Figure 3) or performing a swipe motion (like a windshield wiper) to move the selection marker out of the camera's FOV. In Thumb-over, the selec-tion marker was attached to the cardboard piece and a square piece of paper was attached to the user's thumb. To indicate a selection, the user swiped to occlude the selection marker with the piece of paper. Finally, with Trigger, the selection marker and marker representing the controller were both attached to a make-shift card-board gun with an elastic band trigger. When the trigger was pulled, the selection marker was occluded by stiff paper material. See all marker occlusion techniques in Figure 3.

The pilot study had five participants (3 women, 2 men) aged 25-58 (M = 38 years, SD = 17.6 years) who had at least some to a lot of experience with using handheld tracked controllers. We used a Samsung Galaxy S8 in a cardboard VR viewer with a stereoscopic FOV of 65°. Participants had to complete ISO-9241-9 selection tasks using the three marker occlusion techniques. Condition order was assigned according to a Latin square. Once finished, participants were asked to rank the three techniques in order of preference and to provide verbal comments on the different techniques. Throughout the experiment, the researcher took observational notes on the strategies participants developed while using the controller and selection marker, and what caused them difficulties especially in relation to the tracking technology.

3.3 Pilot Study Results

The pilot study revealed three key themes. First, there were no-ticeable learning effects in using the controller itself. Regardless of condition order, the last condition was significantly easier for partic-ipants to operate the controller and mitigate tracking loss. Second, the main difficulty with the selection marker was finding it with the camera after selection. There was a learning curve to knowing where to hold the selection marker to be in view of the camera. In response to these observations, we updated our main experiment procedure to give participants more time to become comfortable
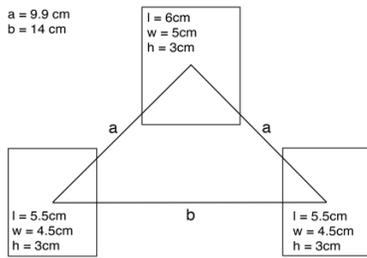
Figure 4: Configuration and measurements of the three multi-target markers used for the prototype. The boxes represent the multi-targets.

with the prototype. Third, the Trigger technique was most difficult due to the marker configuration. The controller marker placed farther up the cardboard was consistently out-of-view of the camera and angling the gun steeply downward blocked the controller marker.

Participants ranked Flip best overall, Thumb-over a close second, and Trigger was ranked worst by every participant except one. Participants stated Flip was easier to operate compared to thumb. The Flip technique provides more freedom in what hand movements a user actuates to hide the marker. This feature is advantageous to potential users with mobility issues. Based on participant feedback, their condition ranking, and observations by researchers, Flip was included into the main experiment below. While a one-handed selection technique is favourable given the decrease in mental and physical workload, the physical design and implementation of Trigger would need a significant overhaul to potentially make it suitable as a form of selection. While this was out-of-scope for the current work, we propose ways to improve Trigger in section 6.1.

3.4 Designing the Controller

We modified the controller design based on the pilot study, replacing the single tracking marker with cubic "multi-target" markers. After testing multiple multi-target sizes and configurations, the final version of the controller design uses three multi-target cubes arranged in an isosceles triangle, as seen in Figure 4.

This marker configuration provided reliable tracking during motion and placing the markers to the lower left and right allowed the controller tracking to operate in more extreme cases, such as being far to the left, right, or forward, effectively expanding the camera FOV. We tested different marker images throughout the design process. The final images forming the multi-target markers were created in Adobe Photoshop and followed Vuforia's best practices for choosing target images such as high contrast, non-repeating patterns, and non-organic shapes.

The virtual controller position is set to the centre of the marker triangle. At least one of three markers must be tracked for the controller to work, and since the distances between each marker is known, the centre can be computed when any marker is tracked. The virtual controller orientation is the average quaternion of the currently tracked markers in a given frame.

We assembled the controller from cardboard, a printer and paper, and tape: materials all commonly found in homes. The only expense to build the controller was the cost of printing the paper to create the

multi-targets. Thus, these design considerations are consistent with our goal to create an accessibly-priced 6DOF tracked controller.

3.5 Designing the Selection Marker

We iterated on the selection marker designed based on the pilot study and further ad hoc testing. We added a cardboard extension to the righthand side of the "handle" to fit a second marker. Since the Cardboard viewer configuration has the smartphone camera is on the user's lefthand side, the selection marker worked more reliably while being held by the left hand. We thus placed the cardboard handle on the left to avoid occluding the camera's view of the controller held by the right hand. Figure 1 (left) depicts the final design of the selection marker apparatus. Separating the selection activation and pointing tasks between two hands mitigates the so-called Heisenberg Effect [8]. The selection marker is only used in the Marker method.

4 METHODOLOGY

We conducted an in-lab user study following the ISO 9241-9 methodology, using a Google Cardboard as the display and our prototype controller as the input device. Our objective was to evaluate the performance and user experience of the Low-Fi VR Controller as a pointing device in mobile VR.

4.1 Participants

We recruited 18 participants (7 women, 10 men, 1 gender fluid) ages 18-44 ($M = 25.7$ years, $SD = 7.8$ years) by posters, email, and through a study recruitment Facebook page. Two participants had no VR experience, but 10 were beginner-to-novice VR users, and six were VR experts. Similarly, 12 participants had prior experience with cardboard VR. All participants had experience using a spatial input device equally claiming very little, moderate, or a lot of experience. All but one were right-handed.

4.2 Apparatus

4.2.1 Hardware. All participants used the POP! CARDBOARD 3.0 by Mr. Cardboard [37] Google Cardboard VR device with a Samsung Galaxy S23 Ultra as the display. The Google Cardboard device has a measured total 70° FOV and stereoscopic FOV of 65° [37]; the FOV may vary depending on eye-to-lens distance and screen size. The device was modified to ensure the view of the smartphone back camera was not obscured, and extra foam padding was added to the nose bridge to ensure comfort throughout the study (see Figure 5). The device included a head strap for hands-free operation.

The Galaxy S23 Ultra ran Android OS 13, and has a 6.8", 1440 × 3088 px (~500 PPI) display with a 120 Hz refresh rate. The back camera has a 24mm focal length and 200MP resolution with $0.6\mu m$ pixel

Figure 5: Google Cardboard Device with front cutout and foam modifications.

Figure 6: Target states from left to right: inactive, active, hover.

Figure 7: The most difficult (A = 200cm, W = 8cm) and least difficult (A = 110cm, W = 16cm) IDs.

size. The smartphone acted as the display, computing device, and its internal sensors were used to detect head motion. Participants could change the viewport orientation with 3DOF head rotation. We used the final version of our Low-Fi VR Controller prototype described in section 3.3 as the pointing device.

4.2.2 Software. We developed a VR system using Unity 2021.3.12f1, the Google Cardboard XR Plugin for Unity v.1.20.0, and the Vuforia Engine AR platform v.10.15.3. The VE used was a room provided by the Google Cardboard XR Plugin. See Figure 7. The VE contained blue, purple, and pink cubes placed around the room and a block texture on the walls to provide additional depth information to the participants. The area of the room was $6.72 \times 6.72 \times 4$ meters, and we positioned the camera 1.7m above the floor. Before each round of selection trials, a start button was visible and floating 2m in front of the user. White text was displayed on the wall in front of participants to provide instructions between selection rounds. We implemented ray-based selection over the virtual hand technique as many studies have found pointing results in more effective selection that requires less physical demand [2, 28, 30]. We did not employ both raycasting and virtual hand techniques as this work is focused on evaluating selection activation methods for the controller. Future work discussed in section 6.1 will explore different selection techniques.

The ISO-9241-9 selection task presents a ring of spherical targets where targets are evenly spaced out and selected one at a time. All targets were placed 3 m in front of participants. The active target participants were instructed to select was coloured dark purple; all other targets were translucent white. When using the Dwell method, the active target's colour smoothly transitioned from purple to white to indicate a selection. With Marker, once the controller ray hovered on the active target (i.e., selection was possible), the target's colour changed to a bright yellow to provide feedback. See Figure 6 for all target states. After a selection was made using any method, there was a two-second delay before the next target was made active. This delay was added to compensate for the time it took for the software to detect the selection marker after being rotated back into view with Marker and was intended to facilitate fair comparison between the conditions. During this delay, the selected target's colour transitioned from red to the inactive state.

The virtual representation of the Low-Fi VR Controller was a 3D model of a PlayStation VR controller from Sketchfab.com[3]. If the controller lost tracking completely, the 3D model turned red to inform the user. A line shooting outward from the controller was continuously rendered to represent the selection ray.

[3]https://sketchfab.com/3d-models/playstation-vr-645bb34486a54e618247c51c98f56205
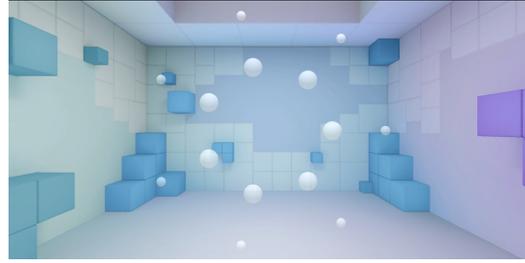
The positions of the three tracked multi-target markers were displayed as red spheres. If a multi-target lost tracking, the corresponding red sphere would disappear. Similarly, the positions of the two image markers used with the Marker method were shown with purple cubes; the cubes disappeared if the images lost tracking. These additional virtual objects were used to signal the current tracking statuses of all the tracked objects to participants. See Figure 1. The software recorded several dependent variables and sent the data to a Google Sheets spreadsheet in real time.

4.3 Procedure

We posted an ad through email and on Facebook advertising user studies to prospective participants. After potential participants reached out and we confirmed their study eligibility, we sent them a consent form and a document describing the controller and study procedure. Participants then scheduled times to come in person to complete the experiment.

During the experiment, participants sat in a swivel chair and wore the Google Cardboard device. They first filled out a demographic survey. Then, the researcher provided oral and written instructions briefing participants on the study objectives and procedure. Once the participant had a clear understanding of what was expected, they began the practice session.

The practice session helped participants familiarize themselves with the controller's tracking behaviour when moving the controller in all directions or using the selection markers. These exercises helped participants get a clear understanding of the spatial, speed, and rotational tracking limitations of the controller. After this, participants performed 72 practice trials (24 per condition). The practice trials included the two easiest and most difficult IDs used in the recorded trials. These two extremes are depicted together in Figure 8. Every participant completed the practice trials in the same order first using the Instant method and ending with Marker.

After the practice session, participants began the recorded trials. The experiment consisted of three selection activation methods, with 12 circles of 7 targets (84 selections total) recorded per condition. Once each target circle was finished, the next circle did not begin until the start button was selected, so participants could take a break. Participants were instructed to select the active targets as quickly and accurately (close to the target centre) as possible. How targets were actually selected depended on the selection activation method, either Instant, Dwell, or Marker. With Instant, targets were immediately selected when touched by the ray. With Dwell, the ray

had to contact the target for at least 300ms to activate selection. We chose 300ms as it was used in similar studies [18, 39, 40] and has been shown to be well-suited to both fast selection and avoiding the Midas Touch problem [4, 18, 19, 35, 51]. With Marker, participants held a cardboard handle with the selection markers attached (see Figure 1) in the smartphone camera's view. To activate selection, participants would point the selection ray at the target, then hide the selection markers from the camera. When both selection markers lost tracking, selection occurred. After a selection, participants would bring the selection markers back in the camera's view to prepare for the next selection.

After completing each condition, participants removed the HMD and filled out the ISO "Independent Questionnaire for assessment of comfort" to assess pointing device comfort [23]. Throughout the experiment, participants were able to take breaks between conditions as desired. After all conditions were completed, participants ranked the selection activation methods in order of preference and provided any feedback on the controller. Participants were then thanked for their time and compensated with $15 CAD.

4.4 Design

The experiment employed a $3 \times 3 \times 2 \times 7 \times 2$ within-subjects design with the following independent variables and levels:

- **Selection Activation:** Instant, Dwell, Marker
- **Amplitude (cm):** 110, 160, 200
- **Width (cm):** 8, 16
- **Trials:** 7
- **Block:** 1, 2

The six combinations of amplitude and width yielded six indices of difficulty (ID): 2.98, 3.46, 3.76, 3.88, 4.39, and 4.7. The most extreme IDs (2.98 and 4.7) are seen in Figure 7. Participants completed two blocks, performing each ID twice. Order of ID was randomized, and the ordering of selection activation was counterbalanced according to a Latin square. Participants completed 36 target circles in total with seven trials each, totaling to 252 total recorded selections each.

There were six dependent variables recorded for each selection: selection time, target re-entries, selection distance, tracking-loss duration, tracking-loss count, and throughput. Selection time was the time in ms from the previous selection to current selection. Target re-entries was the number of times the ray hit the active target after its initial entrance. Selection distance was the distance from the selection point to target centre. Tracking-loss count and tracking-loss duration are the number of times the controller lost tracking, and for how long (in ms) respectively. We calculated and report two variants of throughput, Euclidean-distance and projection-based, both using effective target width and amplitude as described in Equations 3 and 4. Both throughput variants are described in section 2.3.

5 RESULTS

We recorded a total of 4536 selections. Of those, 16 selections timed out at 30s. In the Marker condition, there were 75 selections where participants did not use the selection marker to activate the selection (i.e., it was permanently occluded, so effectively operated like Instant). These 91 data points were removed. In addition, we removed 131 outliers, defined as data points ±3 SDs from the mean
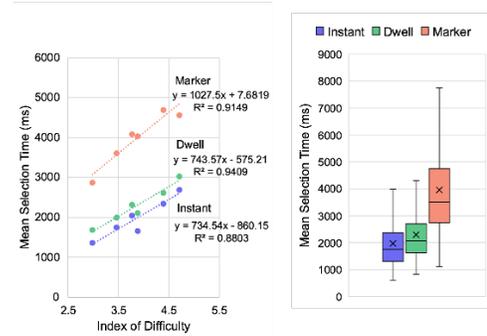
Figure 8: (Left) Fitts' law regression models depicting the relationship between selection time; (Right) Selection time by Selection Activation method.

selection time; most of these were from the Dwell condition. Thus, in total, we removed 212 outliers, corresponding to 4.89% of our data. ANOVA found no significant effect of condition order group on selection time, suggesting counterbalancing was effective.

5.1 Selection Time

Mean selection times for each condition are shown in Figure 8. We performed a Shapiro-Wilk test of normality on the data and found that while the kurtosis parameter had normal-shaped tails, the data set was positively skewed. Therefore, we analyzed selection time using a Friedman test which revealed significant differences between all selection activation methods ($\chi^2 = 32.444$, $p < .001$, $df = 2$). Post hoc pairwise comparisons using Conover's F ($\alpha = .05$) revealed that Marker took significantly longer than Instant and Dwell, while Instant was the fastest. See Figure 8.

We modeled the relationship between selection time and ID and found it to be highly linear with the lowest $R^2 \approx 0.88$ with Instant. See Figure 8. The strong predictive qualities suggest that Fitts' law applies to selection via the Low-Fi VR Controller, and is a good indication the model was accurate despite the task being in 3D.

5.2 Accuracy

Since two of our Selection Activation methods required that a trial end with a successful selection, we do not report traditional error rates (e.g., percentage of targets missed). We instead report timeouts, target re-entries, and selection distance to evaluate accuracy of our conditions. A timeout occurred 30s from the start of the selection, including tracking-loss time. A total of 16 (.35%) of selections timed out; 12 in the Marker condition and 4 in Dwell. Two participants alone yielded 75% of these timeouts. This suggests that Instant was easiest to select with, while Marker was most difficult.

A Shapiro-Wilk test revealed target re-entries were positively skewed. A Friedman test showed target re-entries occurred significantly more frequently with Marker ($\chi^2 = 18.000$, $p < .001$, $df = 1$). Post hoc comparisons showed significant differences between all combinations of width and selection-activation, except Dwell with 0.08cm and Marker with 0.16cm width. See Figure 11. This suggests that it is increasingly difficult to keep the controller steady with smaller targets. This aligns with past work highlighting a limitation
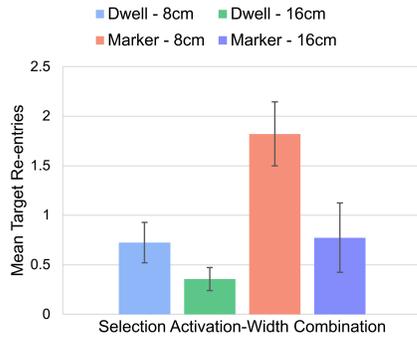
Figure 9: Target re-entries by Selection Activation and Target Width combination.
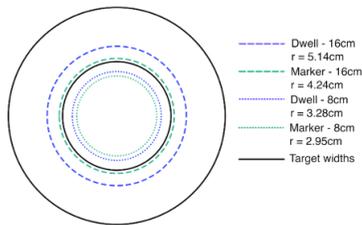
Figure 10: Mean Selection Distance by Selection Activation (excluding Instant) and Target Width conditions. Selection Distance denoted with r.

Figure 11: Euclidean and projection-based throughput by Selection Activation method. Mean throughput score displayed above each data set.
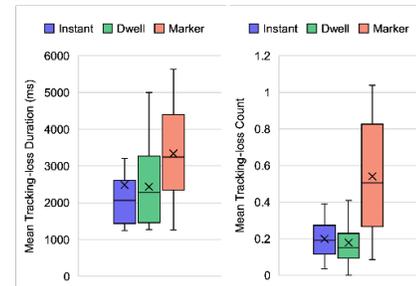
Figure 12: Average Tracking-loss Count per selection (left) and Tracking-loss Duration in ms (right) by Selection Activation method.

of ray-based selection being small and remote targets are difficult to select [2].

Selection distance reveals how close to the target centre the ray was at the time of selection, highlighting control differences between conditions. We exclude Instant from this comparison since the selection distance was always equal to the target radius. A Shapiro-Wilk test revealed the data to be positively skewed. We analyzed the effect of selection activation on selection distance using a Friedman test ($\chi^2 = 14.222$, $p < .001$, $df = 1$). Selections made with Marker were significantly closer to the target centre compared to Dwell. See Figure 12. This was expected, as the selection activation is not automated, so participants could choose exactly when and where to select.

5.3 Throughput

We calculated both the Euclidean-distance throughput and projection-based throughput. Results are seen in Figure 13. A Shapiro-Wilks test revealed both were positively skewed. Friedman tests indicated that the effect of selection activation on both Euclidian-distance throughput ($\chi 2 = 36.000$, p < .001, df = 2) and projection-based throughput ($\chi 2 = 34.111$, p < .001, df = 2) were statistically significant. Post hoc comparisons revealed all selection activation methods were significantly different from each other for both throughput variants. As seen in Figure 13, Instant had a significantly higher throughput, while Marker was significantly lower.

In general, our throughput scores are similar to those reported in past studies using ray-based selection techniques in VR [46, 47, 55].

5.4 Tracking Performance

To assess tracking performance, we looked at both the count of tracking losses and the average duration of tracking losses when they occurred. The mean tracking-loss count per trial for each selection activation method are seen in Figure 12. A Shapiro-Wilk test revealed the tracking-loss counts were positively skewed, so we used a Friedman test, which revealed a significant main effect for selection activation method ($\chi^2 = 25.125$, $p < .001$, $df = 2$). Post hoc comparisons indicated that all selection activation methods had significantly different tracking-loss counts; tracking was lost most often with Marker, and least often with Dwell.

Mean tracking-loss duration for each selection activation method is seen in Figure 12. A Shapiro-Wilk test showed tracking-loss duration was positively skewed; comparison with a Friedman test revealed a significant main effect for selection activation method on tracking-loss duration ($\chi^2 = 24.000$, $p < .001$, $df = 2$). Post hoc comparisons revealed that tracking-loss duration was significantly lower with Instant and Dwell than Marker. Instant and Dwell were not significantly different from each other. We examined a possible correlation between tracking-loss duration and tracking-loss
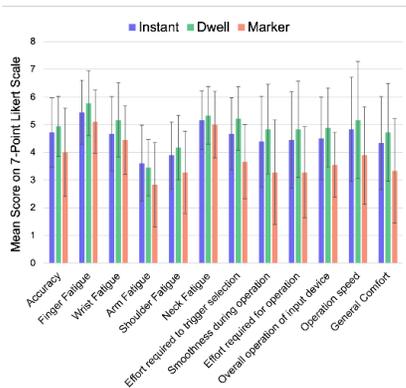
Figure 13: Device Assessment Questionnaire scores by Selection Activation. Error bars show ± SD.

Table 1: Participant ratings of Selection Activation method by preference.

	Instant	Dwell	Marker
1st	9	7	2
2nd	8	9	1
3rd	1	2	15

count by plotting both variables for each ID in each selection activation method in a scatter plot. We calculated the linear regression equation and coefficient of determination for all data combined ($y = 2374.3x - 13.878$, $R^2 = 0.7649$). This yielded a strong positive correlation; regardless of selection activation or task difficulty, as tracking-loss count increased, so did tracking-loss duration.

5.5 User Feedback

After completing each condition, participants rated the given selection activation method using the ISO "Independent Questionnaire for assessment of comfort" [23] where 7 was the best possible score indicating ease of use and low fatigue. As seen in Figure 13, Dwell on average scored highest in all categories except arm fatigue. Meanwhile, Marker on average scored lowest in all categories, especially for statements related to required effort and smoothness of operation.

We also asked participants to rank each selection activation method from most to least preferred. Results are seen in Table 1. Instant and Dwell were by far the most preferred, which aligns with the questionnaire results and participants' objective performance. Instant was overall most preferred.

Participants were also given the opportunity to share any feedback they had on the controller and/or selection activation methods. The main recurring theme among responses involved tracking issues. Once tracking was lost, the controller was slow to detect again. In addition, multiple participants commented on the small FOV of the camera making it difficult to maintain tracking (especially with the Marker selection method). One participant noted the best tracking position was holding the controller higher, but that caused

arm fatigue. Three participants mentioned experiencing jitter, especially when the controller approached the tracking boundary where it became shaky making it difficult to select targets. Finally, two participants stated the Marker selection method was the best concept for real-world use given the extra control it gives the user over selection. However, the tracking issues caused by the selection marker in one hand occluding the controller in the other, made the Marker method most difficult to maintain tracking.

6 DISCUSSION

Overall, the one-handed selection activation methods with automated selection (Instant and Dwell) yielded better selection times (see Figure 9). Despite offering a discrete "click" selection activation method, Marker had considerably worse selection times. These results are consistent with past work [18]. This may in part be due to participants having more difficulty keeping the controller tracked using the Marker method. In particular, there is a visible correlation between selection time and tracking-loss duration, as seen in Figures 8 and 12 With Marker, participants had to keep two objects tracked simultaneously within a small FOV making unintended occlusions likely. This additional mental workload of avoiding these occlusions, plus the bimanual nature of the Marker method likely contributed to the higher tracking-loss durations we observed.

During the experiment, we observed most participants occluding the selection marker with a slow hand movement, which likely increased selection times with Marker. We suspect this more cautious movement was due to unfamiliarity with the device and its less robust materials. Users more experienced with the Low-Fi VR Controller may activate selections significantly faster, perhaps improving selection time with Marker.

It was promising that, in general, selection times using our Low-Fi VR Controller with dwell are comparable to other similar 3D selection techniques [43]. This finding partially addresses R1. Moreover, average throughput scores – regardless of calculation method – are similar to past ray-based selection techniques in VR and other 3D user interfaces. As discussed earlier, a primary merit of throughput is its consistency. Studies reporting throughput scores facilitate comparison between one another due to the reliability of the measure [53]. Computer mouse performance is often viewed as a "gold standard" as it is consistently measured around 4-5 bps [53]. Mouse-based selection typically involves short quick movements in 2D space, so a 3D pointing devices usually offer considerably lower throughput. Past studies report throughput scores ranging from roughly 1.5 bps to around 2.5 bps with similar techniques to our controller and depending on other factors [46, 47, 55]. For example, an HTC Vive controller evaluated in a 3D selection task offered throughput of 1.39 bps [21]. The Low-Fi VR Controller using marker selection activation was slightly lower but comparable at 1.06 and 1.17 bps. Overall, we take this as evidence that under the right circumstances, the Low-Fi VR Controller offers sufficient performance as an input device for MVR, answering R1.

Overall, the best selection times and throughput scores were when using the Instant method, answering R2. We expected Instant to outperform Dwell, as they are very similar selection techniques, but Dwell adds 300ms to the recorded selection time. Despite the

apparent performance benefits of Instant, it is not a realistic technique for real-world applications; it falls victim to the Midas Touch problem [24], as it is "always on".

From our accuracy and error rate results, we conclude that target width has a significant impact on precision. See Figure 11. Target re-entries are likely further affected by "shakiness" and unstable tracking of the controller when reaching the edge of the tracking space, as noted by participants. Despite its poor performance in selection time, the Marker method yielded selections closest to the target centre, reflecting improved accuracy. We speculate this is due to the increased control the Marker technique gives users; the user can choose precisely when and where to select and the two-handed design mitigates potential for the Heisenberg effect of spatial interaction [8].

Fundamental VR interactions the controller can support are grabbing, pointing, travel via teleportation or similar, menu interaction, and potentially basic manipulation tasks if the selection marker is in use. We speculate object translation and rotations could be performed via a click-and-hold action. As is, the controller can potentially be used in games, remote VR research, and educational and collaborative work applications whose main tasks are selection and menu interactions. Future work will focus on enhancing the usability of the controller to help support the described VR scenarios. The following section outlines plans for such improvements.

6.1 Future Work

One goal of this work was to identify potential design improvements for future iterations of the controller. Based on our results, we plan to make three key improvements to the controller: 1) reduce jitter and increase tracking stability, 2) artificially increase the tracking boundaries, and 3) create a one-handed solution that employs Marker selection activation. These three improvements will address R3.

To address the current tracking boundary limits, we plan to update the marker images, their dimensions, the amount, and placements to help keep the controller in the camera FOV under more extreme scenarios such as holding the controller far and low from the camera or angling the controller steeply downwards. This in turn will also improve tracking stability. We note that there is likely an upper limit to adding more markers to improve tracking as too many may hinder users. Therefore, we plan to investigate other solutions such as using a wide-angle camera lens to increase the FOV or developing custom camera-based tracking software to improve overall tracking quality.

We also plan to improve the software by updating how the controller pose is determined to help reduce jitter when the number of tracked markers changes. We also plan to add a lower handle to improve the controller ergonomics and decrease the likelihood of the user's hand blocking a marker.

Finally, although the increased control offered by the Marker technique is preferable for real-world selection tasks, its usability must be improved for it to offer competitive performance. A one-handed version of the Marker selection technique would reduce tracking loss and selection time since occlusion and a small tracking boundary would be less imposing issues. We speculate updating the one-handed version from the pilot study based on our study results could potentially yield performance results comparable to Dwell. Potential updates include changes to the controller markers described above and how the selection marker is occluded. In the pilot study and design phase, we found the cardboard trigger activated too slowly and deteriorated over time. One possible solution is using an LED light concentrated on the selection marker and using a switch to turn on the light to overexpose the marker and hide it from the camera. Future work will compare an updated controller to head-gaze selection (a popular MVR selection technique [45]) and a standard VR 6DOF tracked controller.

6.2 Study Limitations

The two-second delay in between selections (described in section 4.2.2) introduced unintended effects. For multiple participants, the act of holding the controller still on a target was more difficult than the selection task itself due to shakiness. This may have caused increased frustration and arm fatigue over time. Similarly, because each subsequent target was not made active instantly, this introduced a delayed reaction time to the beginning of each recorded selection time; participants had to react to the previously selected target gradually turning white and the next active target turning dark purple. In addition, depending on how far away participants directed their gaze from the centre of the ring of targets, the next activated target may have been out of their field of view, inflating their delayed reaction time and in turn the selection time. For the Marker method, one participant was required to hold the controller in their nondominant hand possibly hindering their performance.

7 CONCLUSION

We proposed the Low-Fi VR Controller: a 6DOF input device for mobile VR that uses a smartphone camera to track markers. This work was motivated by an identified gap in work on improving mobile VR interaction. To our knowledge, there are no previous interaction methods that provide a comparable experience to a high-fidelity VR controller while remaining low cost and accessible.

We evaluated the performance potential and user experience of the controller in an ISO-9241-9 standard task and compared three selection activation methods (Instant, Dwell, Marker). Our results indicate the current design of the controller is a valid input device with comparable selection performance to similar VR input techniques. The controller performs best when operated with one hand and dwell selection activation.

The Low-Fi VR Controller can contribute to the democratization of VR by increasing the complexity of interactions and applications in mobile VR previously only possible with expensive devices. Its highly accessible materials and simple assembly gives great potential for consumer use. It also has potential in supporting remote VR experiments, as the controller can be sent to and assembled by end users, facilitating studies that previously would have required a lab setting. We hope to see and contribute to further developments towards high-complexity, low-cost interactions for mobile VR.

ACKNOWLEDGMENTS

REFERENCES

[1] Mohammed Al-Sada, Shuma Toyama, and Tatsuo Nakajima. 2016. A Mobile VR Input Adaptation Architecture. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 286–287. https://doi.org/10.1145/2994374.3004073

[2] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3: 121–136. https://doi.org/10.1016/j.cag.2012.12.003

[3] Tycho T de Back, Angelica M Tinga, Phong Nguyen, and Max M Louwerse. 2020. Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education* 17: 1–18. https://doi.org/10.1186/s41239-020-00228-9

[4] Per Bækgaard, John Paulin Hansen, Katsumi Minakata, and I Scott MacKenzie. 2019. A fitts' law study of pupil dilations in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 1–5. https://doi.org/10.1145/3314111.3319831

[5] Ravin Balakrishnan and I Scott MacKenzie. 1997. Performance differences in the fingers, wrist, and forearm in computer input control. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 303–310.

[6] Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2019. The Effect of Stereo Display Deficiencies on Virtual Hand Pointing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 1–14. https://doi.org/10.1145/3290605.3300437

[7] Doug A Bowman, Donald B Johnson, and Larry F Hodges. 1999. Testbed Evaluation of Virtual Environment Interaction Techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (VRST '99), 26–33. https://doi.org/10.1145/323663.323667

[8] Doug A Bowman, Chadwick Wingrave, Joshua Campbell, and Vinh Ly. 2001. Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments. Retrieved June 12, 2023 from https://eprints.cs.vt.edu/archive/00000547/01/PinchGlovePaper.pdf

[9] Stuart K Card, Jock D Mackinlay, and George G Robertson. 1991. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems (TOIS)* 9, 2: 99–122.

[10] Mateus C L de Castro, P Joao, Paulo F F Rosa, and Jauvane C de Oliveira. 2021. Interaction by Hand-Tracking in a Virtual Reality Environment. In *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, 895–900. https://doi.org/10.1109/ICIT46573.2021.9453573

[11] Yeonjoo Cha and Rohae Myung. 2013. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics* 43, 4: 350–355. https://doi.org/10.1016/j.ergon.2013.05.005

[12] Bodong Chen. 2015. Exploring the Digital Divide: The Use of Digital Technologies in Ontario Public Schools. *Canadian Journal of Learning and Technology* 41, 3. https://doi.org/10.21432/T2KP6F

[13] Taizhou Chen, Lantian Xu, Xianshan Xu, and Kening Zhu. 2021. Gestonhmd: Enabling Gesture-based Interaction on Low-cost VR Head-Mounted Display. *IEEE Transactions on Visualization and Computer Graphics* 27, 5: 2597–2607. https://doi.org/10.1109/TVCG.2021.3067689

[14] Logan D Clark, Aakash B Bhagat, and Sara L Riggs. 2020. Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies* 139: 102413. https://doi.org/10.1016/j.ijhcs.2020.102413

[15] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6: 381. https://doi.org/10.1037/h0055392

[16] Jan Gugenheimer, David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2016. Facetouch: Enabling touch interaction in display fixed uis for mobile virtual reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, 49–60. https://doi.org/10.1145/2984511.2984576

[17] T Ha and W Woo. 2010. An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, 91–98. https://doi.org/10.1109/3DUI.2010.5444713

[18] 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

[19] Mehedi Hassan, John Magee, and I Scott MacKenzie. 2019. A Fitts' Law Evaluation of Hands-Free and Hands-On Input on a Laptop Computer. In *Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments*, 234–249. http://dx.doi.org/10.1007/978-3-030-23563-5_20

[20] Teresa Hirzle, Jan Rixen, Jan Gugenheimer, and Enrico Rukzio. 2018. Watchvr: Exploring the usage of a smartwatch for interaction in mobile virtual reality. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3188629

[21] Wen-jun Hou and Xiao-lin Chen. 2021. Comparison of Eye-Based and Controller-Based Selection in Virtual Reality. *International Journal of Human–Computer Interaction* 37, 5: 484–495. https://doi.org/10.1080/10447318.2020.1826190

[22] Akira Ishii, Takuya Adachi, Keigo Shima, Shuta Nakamae, Buntarou Shizuki, and Shin Takahashi. 2017. FistPointer: target selection technique using mid-air interaction for mobile VR environment. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 474. https://doi.org/10.1145/3027063.3049795

[23] ISO. 2000. ISO 9241-9:2000 Ergonomic requirements for office work with visual display terminals (VDTs) — Part 9: Requirements for non-keyboard input devices. International Organization for Standardization.

[24] Robert J K Jacob. 1990. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '90), 11–18. https://doi.org/10.1145/97243.97246

[25] Daniel Kharlamov, Brandon Woodard, Liudmila Tahai, and Krzysztof Pietroszek. 2016. TickTockRay: smartwatch-based 3D pointing for smartphone-based virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 365–366. https://doi.org/10.1145/2993369.2996311

[26] Werner A König, Jens Gerken, Stefan Dierdorf, and Harald Reiterer. 2009. Adaptive pointing–design and evaluation of a precision enhancing technique for absolute pointing devices. In *Human-Computer Interaction–INTERACT 2009: 12th IFIP TC 13 International Conference, Uppsala, Sweden, August 24-28, 2009, Proceedings, Part I 12*, 658–671. https://doi.org/10.1007/978-3-642-03655-2_73

[27] Stanislav Kyian and Robert Teather. 2021. Selection Performance Using a Smartphone in VR with Redirected Input. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, 1–12. https://doi.org/10.1145/3485279.3485292

[28] Joseph J. LaViola, Ernst Kruijff, Doug A. Bowman, and Ivan Poupyrev. 2017. Selection and Manipulation. In *3D User Interfaces* (2nd ed.). Addison-Wesley Professional.

[29] Richard Li, Victor Chen, Gabriel Reyes, and Thad Starner. 2018. ScratchVR: low-cost, calibration-free sensing for tactile input on mobile virtual reality enclosures. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, 176–179. https://doi.org/10.1145/3267242.3267260

[30] Siqi Luo, Robert J Teather, and Victoria McArthur. 2020. Camera-based selection with cardboard head-mounted displays. In *International Conference on Human-Computer Interaction*, 383–402. https://doi.org/10.1007/978-3-030-59990-4_29

[31] Xiaolei Lv, Chengqi Xue, and Weiye Xiao. 2021. Influence of Size and Depth Perception on Ray-Casting Interaction in Virtual Reality. In *Intelligent Human Systems Integration 2021*, 152–158. https://doi.org/10.1007/978-3-030-68017-6_23

[32] I. Scott MacKenzie. 2018. *The Wiley Handbook of Human Computer Interaction*. Wiley. https://doi.org/10.1002/9781118976005

[33] I Scott MacKenzie and Poika Isokoski. 2008. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1633–1636. https://doi.org/10.1145/1357054.1357308

[34] I Scott MacKenzie and Robert J Teather. 2012. FittsTilt: The application of Fitts' law to tilt-based interaction. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, 568–577. https://doi.org/10.1145/2399016.2399103

[35] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 357–360. https://doi.org/10.1145/1518701.1518758

[36] Peter Mohr, Markus Tatzgern, Tobias Langlotz, Andreas Lang, Dieter Schmalstieg, and Denis Kalkofen. 2019. TrackCap: Enabling Smartphones for 3D Interaction on Mobile Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3290605.3300815

[37] MR. Cardboard. 2021. POP! CARDBOARD 3.0 – Inspired by Google Cardboard. *Mr. Cardboard*. Retrieved from https://mrcardboard.eu/product/pop-cardboard-2-0-made-in-germany-inspired-by-google-cardboard/

[38] Atsuo Murata and Hirokazu Iwase. 2001. Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science* 20, 6: 791–805. https://doi.org/10.1016/S0167-9457(01)00058-6

[39] Aunnoy Mutasim, Anil Ufuk Batmaz, Moaaz Hudhud Mughrabi, and Wolfgang Stuerzlinger. 2022. Performance Analysis of Saccades for Primary and Confirmatory Target Selection. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, 1–12. https://doi.org/10.1145/3562939.3565619

[40] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (ETRA '21 Short Papers). https://doi.org/10.1145/3448018.3457998

[41] J Logan Olson, David M Krum, Evan A Suma, and Mark Bolas. 2011. A design for a smartphone-based head mounted display. In *2011 IEEE Virtual Reality Conference*, 233–234. Retrieved June 28, 2023 from https://www.markbolas.com/_files/ugd/60edd5_9ca15584daff4c3a91aa07f12fd22e5d.pdf

[42] Pew Research Center. 2021. Mobile Fact Sheet. *Pew Research Center*. Retrieved December 29, 2021 from https://www.pewresearch.org/internet/fact-sheet/mobile/

[43] Duc-Minh Pham and Wolfgang Stuerzlinger. 2019. Is the Pen Mightier than the Controller? A Comparison of Input Devices for Selection in Virtual and Augmented Reality. In *25th ACM Symposium on Virtual Reality Software and Technology* (VRST '19). https://doi.org/10.1145/3359996.3364264

[44] Krzysztof Pietroszek, Liudmila Tahai, James R Wallace, and Edward Lank. 2017. Watchcasting: Freehand 3D interaction with off-the-shelf smartwatch. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 172–175. https://doi.org/10.1109/3DUI.2017.7893335

[45] Wendy Powell, Vaughan Powell, Phillip Brown, Marc Cook, and Jahangir Uddin. 2016. Getting around in google cardboard–exploring navigation preferences with low-cost mobile VR. In *2016 IEEE 2nd Workshop on Everyday Virtual Reality (WEVR)*, 5–8. https://doi.org/10.1109/WEVR.2016.7859536

[46] Yuan Yuan Qian and Robert 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 5th Symposium on Spatial User Interaction*, 91–98. https://doi.org/10.1145/3131277.3132182

[47] Adrian Ramcharitar and Robert J Teather. 2018. EZCursorVR: 2D Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 44th Graphics Interface Conference*, 123–130. https://doi.org/10.20380/GI2018.17

[48] Donald C Reitzes, Josie Parker, Timothy Crimmins, and Erin E Ruel. 2016. Digital communications among homeless people: Anomaly or necessity? *Journal of Urban Affairs*.

[49] Gang Ren, Xiao Liu, and Jingfu Liu. 2019. Design and Evaluation of 3D Selection in Mobile VR Environments. In *2019 8th International Conference on Innovation, Communication and Engineering (ICICE)*, 17–20. https://doi.org/10.1109/ICICE49024.2019.9117494

[50] Albert Skip Rizzo, Belinda Lange, Evan A Suma, and Mark Bolas. 2011. Virtual Reality and Interactive Digital Game Technology: New Tools to Address Obesity and Diabetes. *Journal of Diabetes Science and Technology* 5, 2: 256–264. https://doi.org/10.1177/193229681100500209

[51] Asma Shakil, Christof Lutteroth, and Gerald Weber. 2019. Codegazer: Making Code Navigation Easy and Natural with Gaze Input. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300306

[52] William J Shelstad, Dustin C Smith, and Barbara S Chaparro. 2017. Gaming on the rift: How Virtual Reality Affects Game User Satisfaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2072–2076. https://doi.org/10.1177/1541931213602001

[53] R William Soukoreff and I Scott MacKenzie. 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. https://doi.org/10.1016/j.ijhcs.2004.09.001

[54] Robert J Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, 87–94. https://doi.org/10.1109/3DUI.2011.5759222

[55] Robert J Teather and Wolfgang Stuerzlinger. 2013. Pointing at 3D Target Projections with One-Eyed and Stereo Cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 159–168. https://doi.org/10.1145/2470654.2470677

[56] Wen-Jie Tseng, Li-Yang Wang, and Liwei Chan. 2019. FaceWidgets: Exploring Tangible Interaction on Face with Head-Mounted Displays. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 417–427. https://doi.org/10.1145/3332165.3347946

[57] Aditya Vishwanath, Matthew Kam, and Neha Kumar. 2017. Examining Low-Cost Virtual Reality for Learning in Low-Resource Environments. In *Proceedings of the 2017 conference on designing interactive systems*, 1277–1281. https://doi.org/10.1145/3064663.3064696

[58] Jeenal Vora, Santosh Nair, Anand K Gramopadhye, Andrew T Duchowski, Brian J Melloy, and Barbara Kanki. 2002. Using virtual reality technology for aircraft visual inspection training: presence and comparison studies. *Applied Ergonomics* 33, 6: 559–570. https://doi.org/10.1016/S0003-6870(02)00039-X

[59] B K Wiederhold, D P Jang, R G Gevirtz, S I Kim, I Y Kim, and M D Wiederhold. 2002. The treatment of fear of flying: a controlled study of imaginal and virtual reality graded exposure therapy. *IEEE Transactions on Information Technology in Biomedicine* 6, 3: 218–223. https://doi.org/10.1109/TITB.2002.802378

[60] Xiaoqi Yan, Chi-Wing Fu, Pallavi Mohan, and Wooi Boon Goh. 2016. CardboardSense: Interacting with DIY Cardboard VR Headset by Tapping. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 229–233. https://doi.org/10.1145/2901790.2901813

[61] L Ying, Z Jiong, S Wei, W Jingchun, and G Xiaopeng. 2017. VREX: Virtual reality education expansion could help to improve the class experience (VREX platform and community for VR based education). In *2017 IEEE Frontiers in Education Conference (FIE)*, 1–5. https://doi.org/10.1109/FIE.2017.8190660

[62] Soojeong Yoo and Callum Parker. 2015. Controller-less Interaction Methods for Google Cardboard. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, 127. https://doi.org/10.1145/2788940.2794359

[63] Difeng Yu, Weiwei Jiang, Chaofan Wang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. ShadowDancXR: Body Gesture Digitization for Low-cost Extended Reality (XR) Headsets. In *Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces*, 79–80. https://doi.org/10.1145/3380867.3426222

[64] Shumin Zhai, Paul Milgram, and William Buxton. 1996. The influence of muscle groups on performance of multiple degree-of-freedom input. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 308–315. https://doi.org/10.1145/238386.238534

[65] L Zheng, T Xie, and G Liu. 2018. Affordances of Virtual Reality for Collaborative Learning. In *2018 International Joint Conference on Information, Media and Engineering (ICIME)*, 6–10. https://doi.org/10.1109/ICIME.2018.00011