# Cardboard Controller: A Cost-Effective Method to Support Complex Interactions in Mobile VR

Kristen Grinyer and Robert J. Teather

***Abstract*—To address the need for high-complexity low-cost interaction methods for mobile VR, we present a Cardboard Controller, supporting 6-degree-of-freedom target selection while being made of low-cost, highly accessible materials. We present two studies, one evaluating selection activation methods, and the other comparing performance and user experience of the Cardboard Controller using ray-casting and the virtual hand. Our Cardboard Controller has comparable throughput and task completion time to similar 3D input devices and can effectively support pointing and grabbing interactions, particularly when objects are within reach. We propose guidelines for designing low-cost interaction methods and input devices for mobile VR to encourage future research towards the democratization of VR.**

***Index Terms*—Virtual reality, interaction techniques, input devices.**

## I. Introduction

VIRTUAL reality (VR) head-mounted displays (HMDs) are now advertised for every-day use for work, socializing, entertainment, and personal fitness. Extended reality (XR) technologies including VR are now becoming the next major computing paradigm like the smartphone and desktop before it. This paradigm shift threatens to widen the digital divide, the gap between those *with and* without access to technologies [15]. Most VR HMDs cost between $500-1000 USD, but can reach upwards of $3500 USD [8], prohibitively expensive for people of lower socio-economic status (SES) to use the technology. With mass VR adoption on the horizon, VR device design must consider equitable access.

Mobile VR (MVR) provides a low-cost alternative using a smartphone for both display and computing. The smartphone is inserted into a cardboard viewer costing $10-15 USD. Although MVR assumes the user owns a smartphone, a 2024 survey of 5626 U.S. adults found that 91% owned smartphones; 84% of adults with an income under $30k USD and 85% of adults with an education of high school or less owned a smartphone [31]. Smartphones mass adoption enables MVR to improve VR access across demographics. However, MVR currently does not support effective 3D interaction needed in most VR applications. MVR supports rotational head tracking only; most VR applications require hand- or controller-based interaction to support fundamental VR tasks (i.e., selection, object manipulation, and travel). To date, MVR cannot support the most common VR interaction techniques (i.e., virtual hands and ray-casting [1], [24]), so cannot effectively address the digital divide presented by VR technologies.

Fig. 1. (Left) Virtual controller with selection ray; four floating shapes represent controller and selection apparatus tracked markers.(Right) Cardboard Controller in right hand and selection apparatus in left.

We propose the Cardboard Controller (see Fig. 1), a 6-degree-of-freedom (DOF) input device made of paper materials. It offers 6DOF interaction but is accessible across SES due to its low cost. The smartphone's rear-facing camera tracks three paper markers attached to the cardboard handle to facilitate controller pose tracking. We also note the need for a "click" action to indicate selection. Previous work on MVR interaction evaluated low-cost selection activation [5], [23], [40] and hand tracking [6], [24]. However, prior work used extra, and expensive, external devices such as a smartwatch or a second smartphone for clicking [14], [21], [27].

We evaluated user performance offered by the Cardboard Controller in two experiments conforming to the ISO 9241-411 [17] standard. The first experiment exclusively used ray-based selection and compared three selection activation methods including our novel 'Marker' method. The second experiment compared three selection techniques: virtual hand, ray-casting, and head-gaze (the most common selection technique in MVR [34]). We aim to improve the controllers' performance while answering the following research questions:

- R1 Is the proposed Marker selection activation effective for manual selection activation?
- R2 What performance (selection time and throughput) does the Cardboard Controller offer with ray-casting and virtual hand?
- R3 Which selection technique is best suited to the Cardboard Controller in terms of performance and user opinion?

The contributions of our work include:

Received 18 October 2024; revised 14 June 2025; accepted 15 June 2025. Date of publication 19 June 2025; date of current version 5 September 2025. This work was supported in part by the Natural Science and Engineering Research Council of Canada (NSERC). Recommended for acceptance by J. Grubert. *(Corresponding author: Robert J. Teather.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Carleton U Research Ethics Board-B under Application No. 115464.

The authors are with Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: kristengrinyer@cmail.carleton.ca; rob.teather@carleton.ca).

Digital Object Identifier 10.1109/TVCG.2025.3581158

- The design of our Cardboard Controller and two experiments providing evidence that the controller offers reasonable performance in near and remote selection tasks using common VR selection techniques.
- Formal comparison of head-gaze-based selection to ray-casting and virtual hand selection techniques.
- Design guidelines for low-cost input devices and interaction techniques for MVR that can be used to further develop complex interactions in MVR.

## II. Related Work

We developed the Cardboard Controller for object selection, so review past approaches to support selection in MVR and evaluate user experience (UX) of VR input devices.

### A. User Experience of VR Systems

VR offers an extensive range of possible interaction techniques and novel input devices, thus it is important to examine how these factors influence the UX of VR systems. Notably, in their systematic literature review, Kim et al. [20] report a multitude of factors affect VR system UX including attributes of the user, devices, task, and environment. Since VR systems frequently have the user standing, they typically employ tracked handheld controllers to support the use of natural motor abilities. These devices afford a more natural UI than other devices such as gamepad controllers, mouse, or keyboard; this is known to improve perceived immersion and increase spatial presence [16], [20], [26], as well as offer more enjoyable gaming experiences in VR [26]. However, Hufnal et al. [16] found no significant difference in user satisfaction between a tracked VR controller and a gamepad controller. Compared to hands-free interaction, handheld devices can support vibrotactile haptic feedback, which has enhances the UX of VR [38]. However, handheld devices add encumbrance; this can decrease physical comfort and tracking reliability. As such, when designing VR handheld input devices, one should also consider ergonomic factors [20]. We applied this principle when collecting user feedback during both of our experiments.

### B. Selection in VR

After locomotion, selection is the most common fundamental interaction task in VR and commonly precedes other tasks (e.g., manipulation, system control) [1]. The quality and appropriateness of a selection technique impacts user performance and experience; a poorly designed or unsuitable technique hinders overall performance. Selection techniques are often classified by metaphor [1], [4], [33]; the most common are the virtual hand and pointing metaphors [1], [24]. With virtual hands, the user holds a controller or tracker, which controls a virtual hand proxy using a 1-1 or offset mapping for direct manipulation [1] providing natural and intuitive interaction [4]. Virtual hands are useful for high-precision tasks.

Virtual hands are limited to in-reach targets, but remote pointing techniques address this issue; users control the direction and origin of a ray, or a volume such as a cylinder or cone, and select objects intersected by it [1]. The ray usually originates from an input device or the user's head/hand position. Remote pointing offers better selection performance than other selection metaphors, including the virtual hand [1].

### C. Improving MVR Selection

Without 6DOF tracking, MVR cannot support interaction techniques required to leverage the benefits of high-fidelity VR devices [12], [19], [23], [24], [34]. Selection is most commonly performed with head-gaze pointing [19], where the ray originates at the head and its direction is controlled by the phone's orientation. Selection activation occurs either using dwell (i.e., a brief timeout when the cursor rests on a target) or a lever on the cardboard HMD [19]. Dwell causes accidental selections [19] and inflates selection time. However, the Cardboard lever is unreliable and susceptible to the "Heisenberg Effect" [19]: unintentional device movement at the instant of selection, resulting in error.

Past MVR selection work focused on three main challenges: 1) input is limited to the lever or dwell, 2) tracking is limited to 3DOF head rotation, and 3) the smartphone's limited processing power and field of view (FOV). Compared to high-end VR, proposed solutions offer limited performance, or require additional expensive devices contrary to the low-cost nature of MVR. We describe various approaches below.

*1) Additional Hardware Solutions:* Past solutions used inertial sensors in a smartwatch [14], [19] for 3DOF control of a ray originating from the HMD. Selection activation used either the watch face as a button [14] or forearm rotation [19]. Other approaches optically tracked a second smartphone as a handheld controller [21], [27]. This approach supports both virtual hand and pointing selection, 6DOF object manipulation, and 2D input on the smartphone screen. These solutions offer the best MVR interaction, but the highest cost due to the need for a second smartphone for an input device.

Other work integrated hand tracking [6], [29] either through a Leap Motion sensor mounted on the wrist and connected to the smartphone [29] or by capturing video from the smartphone camera and processing it on a nearby PC [6]. While these approaches offer 6DOF virtual hand interaction, they are computationally expensive and require secondary processing the smartphone was found to be insufficient for [29].

All of the aforementioned solutions impose additional costs, contradicting the goal of MVR: to be portable and accessible to users of any SES [34]. We aim to support complex interactions without the need of additional expensive equipment.

*2) Low-Cost Solutions:* Other work proposed solutions requiring little to no extra cost, either focused on improving MVR tracking [7], [9], [24] or using HMD surfaces for interaction [5], [23], [40]. Luo et al. [24] used the smartphone camera alone to track the user's hand and fingers supporting virtual hand and fixed origin ray-casting, but found performance was very low. Eye tracking has also been considered [7]. The method used the smartphone's front camera to track users' pupils, required no extra hardware or heavy processing, and offered comparable accuracy to eye trackers in modern VR HMDs in the central FOV of $\sim 20°$of visual angle.

Others have used the surfaces of Google Cardboard for UI interaction and input [5], [23], [40], for example, using machine learning to detect tap and sliding surface gestures [5], [40].

Another approach used a magnet and washer pushed around a circular cardboard track for menu scrolling [23].

While these solutions are promising, our Cardboard Controller is a low-cost tangible input device, necessary to support the majority of VR applications made for higher-end HMDs.

### D. Fitts' Law

Our evaluation employs the ISO 9241-411 standard based on Fitts' law, which models rapid aimed movements such as selection tasks [25]. 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*). *ID* is measured in bits, and a and b are the intercept and slope of the linear regression line illustrating this relationship.

$$MT = a + bID \tag{1}$$

$$ID = log_2(A/W + 1) \tag{2}$$

Throughput (*TP*) combines selection speed and accuracy and is commonly used to quantify human performance. It has been shown to be consistent despite speed/accuracy trade-offs [25]. This property, along with its consistency between studies make it the preferred option for comparing selection techniques [36]. It is independent of *A* and *W*, since as *ID* changes, *MT* changes inversely [25]. *TP* is defined as:

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

$$ID_e = log_2(A_e/W_e + 1) \tag{4}$$

$ID_e$ is the effective index of difficulty (4); the effective measures adjust $W$ and $A$ to make *TP* more consistent across studies [21], [37]. Effective width, $W_e = 4.13 \times SD_x$, uses the standard deviation, $SD_x$, of the selection coordinate over/undershooting along the task axis (line between subsequent targets). Assuming selection coordinates are normally distributed around the target centre, $\pm 2.066$ (or 4.133) standard deviations corresponds to 96% of selections hitting the target. This so-called accuracy adjustment sets the experimental error rate to 4% [25] facilitating *TP* comparison across studies with varying error rates [36]. Effective amplitude, $A_e$, is the average distance the cursor moves for each selection. Both effective measures better represent participant performance.

The ISO 9241-411 standard [17] recommends *TP* as a main dependent variable. We employ this metric in our experiment and report two variations. One projects the closest point on the selection ray onto the task axis (line between subsequent targets) effectively providing a 1D $SD_x$ value. The other uses the euclidean 3D distance from the target centre to the selection coordinate, which penalizes inaccuracy in depth [37].

## III. Controller Design Process

### A. Initial Design

We devised several Cardboard Controller prototypes and informally tested them and various software/hardware setups. Ultimately, we used Vuforia v.10.15.3 and the Google XR Plugin v.1.2 in Unity 2021. We compared several marker configurations and selection activation methods in prototyping. Initial prototypes used one tracked marker with 300 ms dwell for selection activation. We chose 300 ms as it is well-suited for fast selection and avoiding unintentional selections [13].

Since manual selection activation (e.g., a button press) is preferable to automatic methods (e.g., dwell), we devised a manual selection activation method dubbed *selection marker*, which used a second tracked marker that activated selection when occluded from the camera view. A pilot study [9] revealed that holding the selection marker in the non-dominant hand and flipping it out of the camera FOV worked best (see Fig. 1). Separating pointing and selection activation between hands also mitigates the Heisenberg Effect [3].

### B. Final Design

Based on the pilot study, We replaced the single marker with three "multi-target" cubic markers arranged in an isosceles triangle. This marker configuration provided reliable tracking during motion and increased controller visibility at the edges of the camera FOV. The tracked images used were created in Adobe Photoshop and followed Vuforia's best practices for target images. The virtual controller position is the centre of the marker triangle. At least one marker must be tracked to compute the centre. In experiment 1 (E1), the controller orientation was the average quaternion of the tracked markers. This was modified in experiment 2 (E2) to use a single tracked marker quaternion to reduce jitter. By default, this was the central marker, unless it was not tracked in the previous frame, in which case another visible marker's quaternion was used. We assembled the controller from cardboard, paper, and tape. These are common household materials consistent with our goal to create an accessibly-priced 6DOF controller. Fig. 1 depicts the final design.

## IV. Experiment 1

We conducted a study using the ISO 9241-411 methodology, to determine ray-casting performance of the controller with different selection activation methods. The study was approved by our university ethics board.

### A. Participants

We recruited 18 participants (7 women, 10 men, 1 gender fluid) aged 18-44 ($M = 25.7$ years, $SD = 7.8$ years) by posters, email, and Facebook. Two participants had no VR experience, but 10 were novice VR users, and six were VR experts. Similarly, 12 participants had prior experience with cardboard VR. All participants had at least some experience with spatial input devices. All but one were right-handed.

### B. Apparatus

*1) Hardware:* We used the POP! CARDBOARD 3.0 by Mr. Cardboard [28] with a Samsung Galaxy S23 Ultra. The Google Cardboard has a stereoscopic FOV of 65° [28]. We modified the Cardboard to expose the smartphone's rear camera and added extra foam padding to the nose bridge for comfort, and added a strap for hands-free operation (see Fig. 2).

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 rear camera has a 24 mm focal length and 200 MP resolution. The smartphone acted as the display, computing device, and its

Fig. 2. Google Cardboard Device with front cut-out to expose middle camera, blue foam on face and nose bridge, and head strap.

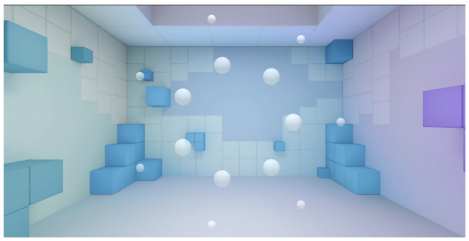


Fig. 3. Two target circles are overlaid in VE. Outer circle shows highest *ID* (4.39) with smallest target width and largest amplitude. Inner circle shows smallest *ID* (2.98) with largest target width and smallest amplitude.

internal sensors were used to detect head motion (e.g., viewport orientation was controlled by 3DOF head rotation). We used the final version of our Cardboard Controller prototype described in Section III-B.

*2) Software:* We used Unity 2021.3, the Google Cardboard XR Plugin v.1.20.0, and the Vuforia Engine AR platform v.10.15.3. The VE was a room provided by the Google Cardboard XR Plugin. See Fig. 3. The room measured 6.72 × 6.72 × 4 m, and we positioned the camera 1.7 m above the floor. Instructions were periodically provided by white text.

We used the ISO-9241-411 selection task, which presents a ring of spherical targets evenly spaced and selected sequentially. All targets were placed 3 m in front of participants. Participants used ray-casting to select the active target which was coloured dark purple; all other targets were translucent white. When using the Dwell selection activation method, the active target's colour smoothly transitioned from purple to white to indicate selection. With Marker, once the controller ray hovered on the active target (i.e., selection was possible), the target's colour changed bright yellow. If 30 seconds passed, the selection was considered timed-out and the next target was made active. After selection, there was a two-second delay before the next target activated, to compensate for the time required to detect the selection marker after being rotated back into view with Marker. During this delay, the selected target's colour transitioned from red to the inactive state.

The virtual controller was a 3D model of a PlayStation VR controller from Sketchfab.com. If the controller lost tracking, the model turned red. The selection ray was continuously rendered as a line. The three tracked markers were displayed as red spheres, and would disappear if tracking was lost on the corresponding marker. The selection markers used with the Marker method were shown as purple cubes, which also disappeared if they lost tracking. These additional objects gave feedback on the tracking status of all tracked objects. See Fig. 1. The software recorded several measurements and sent the data to a Google Sheets spreadsheet in real time.

### C. Procedure

Upon arrival, after providing informed consent, participants sat in a swivel chair and wore the Google Cardboard device. They first filled out a demographic survey. Then, we provided oral and written instructions. Next, they completed a practice session to familiarize them with the controller's tracking behaviour, and to understand the spatial, speed, and rotational tracking limitations of the controller. They next performed 24 practice trials per condition (72 total) using the two easiest and most difficult *ID*s. See Fig. 3. Practice trials always used the same condition order (Instant, Dwell, then Marker).

After the practice session, participants began the recorded trials, consisting of three selection activation methods, with 12 circles of 7 targets (84 selections) recorded per condition. Participants could take breaks after completing each target circle. Participants were instructed to select the targets as quickly and accurately as possible.

After completing each condition, participants removed the HMD and filled out the ISO questionnaire to assess device comfort [17]. Throughout the experiment, participants could take breaks between conditions. After completing all conditions, participants ranked the selection activation methods by preference, and provided any feedback on the controller. They were compensated with $15 CAD upon completion.

### D. Design

We employed a 3 × 3 × 2 × 7 × 2 within-subjects design with the following independent variables and levels:

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

Selection activation methods include Instant (immediate selection when the ray touches the target), Dwell (300 ms of hovering the ray on targets to select), and Marker (using the selection markers described in Section III-B). Differences in visual feedback between selection activation methods are described in Section IV-B2. The six combinations of amplitude and width yielded six *ID*s: 2.98, 3.46, 3.76, 3.88, 4.39, and 4.7. The most extreme *ID*s (2.98 and 4.7) are seen in Fig. 3. Participants completed two blocks, performing each *ID* twice. *ID* order was randomized, and selection activation order was counterbalanced with a Latin square. Participants completed 36 target circles in total with seven trials each, for 252 total selections. There were six dependent variables recorded for each selection: selection time (ms), target re-entries (count), selection distance (% from target centre), tracking-loss duration (ms), tracking-loss count, and throughput (bps).

### E. Results

We recorded a total of 4536 trials. In the Marker condition, we removed 75 trials where participants permanently occluded the selection marker, so it operated like Instant. We further removed 131 outliers, data points ± 3 *SD*s from the mean selection time; most were with Dwell. In total, we removed 222 outliers (4.89% of our data). We performed Shapiro-Wilk normality tests on all

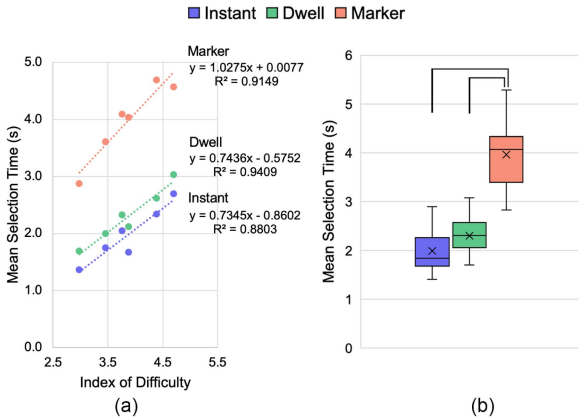



Fig. 4. (a) Fitts' law regression models depicting relationship between selection time and *ID*; (b) Selection time by Selection Activation. Sig. diff. of $p <$. 001 shown.

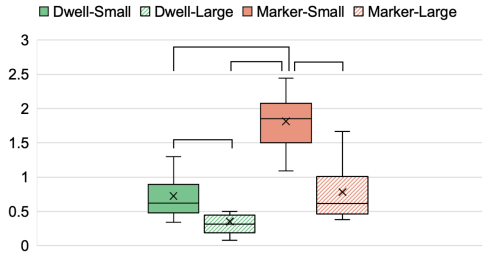



Fig. 5. Target re-entries by Selection Activation and Target Width combination. Sig. diff. of $p <$. 001. shown.

data sets. All were positively skewed, so our analysis uses Friedman tests. All post-hoc analyses used Conover's F ($\alpha =$. 05). To compute statistical power, we followed the recommendation of Lehmann [22] to first compute the power of the parametric equivalent (ANOVA), then subtract 15% as an adjustment. We also report effect size, represented using Kendall's *W*, for the Friedman tests.

*1) Selection Time:* Selection time was the time in ms from the previous to current selection. Mean selection times for each condition are seen in Fig. 4. A Friedman test revealed a significant main effect for selection activation ($\chi^2 = 32.44, p <$. 001, $df = 2$, $W =$. 9, *pow.* =. 91). All pairwise significant results found by posthocs are seen in the figures.

We modelled the relationship between selection time and *ID* and found it highly linear (lowest $R^2 \approx 0.88$ with Instant). See Fig. 4. The strong predictive qualities shows that Fitts' law applies to selection with the Cardboard Controller, and the model was accurate despite the task being in 3D.

*2) Accuracy:* Since two selection activation methods required trials end with a successful selection, we do not report error rates (e.g., % of targets missed). We instead report target re-entries and selection distance to assess accuracy. Re-entries was the number of times the ray hit the target after the initial hit, while selection distance was the euclidean distance between the target centre and the closest point on the ray at the time of selection. A Friedman test showed target re-entries occurred significantly more frequently with Marker ($\chi^2 = 18.0, p <$. 001, $df = 1$, $W = 1.0$, *pow.* =. 51). See Fig. 5.

We present selection distance as the percentage from target centre, where 0% is perfectly on centre while 100% is

TABLE I
AVERAGE PERCENTAGE AWAY FROM TARGET CENTRE

| Selection Activation | Dwell | Dwell | Marker | Marker |
|---|---|---|---|---|
| Target Width | small | large | small | large |
| % from Target Centre | 41% | 32.2% | 37.1% | 26.6% |

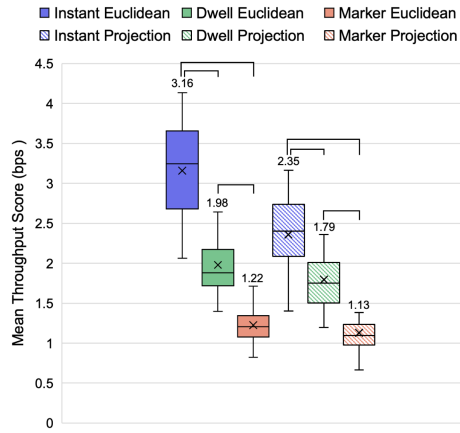



Fig. 6. Euclidean and projection-based *TP* by Selection Activation. Mean *TP* score displayed above each data set. Sig. diff. of $p <$. 001 shown.

on the target sphere surface. We exclude Instant here, since selection distance with Instant was always equal to target radius. A Mann Whitney U-test found a significant difference between Marker and Dwell ($z =$ -3.23, $p <$. 05, $r =$ .54, *pow.* =. 93). Selections with Marker were significantly closer to centre than Dwell, perhaps because participants could choose exactly when and where to select, similar to a button. See Table I.

*3) Throughput:* We calculated and report two variants of throughput, euclidean-distance and projection-based, both using effective target width and amplitude as described in (3) and (4). Both *TP* variants are described in Section II-D. Results are seen in Fig. 6. Friedman tests indicated that the effect of selection activation on both Euclidian-distance *TP* ($\chi^2 = 36.0, p <$. 001, $df = 2$, $W = 1.0$, *pow.* =. 51) and projection-based *TP* ($\chi^2 = 36.0, p <$. 001, $df = 2$, $W = 1.0$, *pow.* =. 74) were statistically significant. Fig. 6 depicts *TP* scores and significant pairwise differences. In general, our *TP* scores are similar to those reported in past studies using ray-based selection techniques in VR [35], [37].

*4) Tracking Performance:* To assess tracking performance, we looked at the number of times the controller lost tracking (tracking-loss count) and for how long in ms (tracking-loss duration). The mean tracking-loss count per trial for each selection activation method are seen in Fig. 7. A Friedman test revealed a significant main effect for selection activation method ($\chi^2 = 25.13, p <$. 001, $df = 2$, $W =$. 7, *pow.* =. 53).

Mean tracking-loss duration for each selection activation method is seen in Fig. 7. A Friedman test revealed a significant main effect for selection activation method on tracking-loss duration ($\chi^2 = 24.0$, $p <$. 001, $df = 2$, $W =$. 67, *pow.* =. 09). We examined a possible correlation between tracking-loss duration and tracking-loss 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 - 13.88,

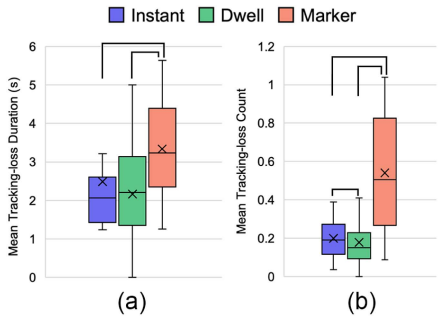



Fig. 7. (a) Mean Tracking-loss Count and (b) Tracking-loss Duration per selection by Selection Activation. Sig. diff. of $p < .001$ shown.

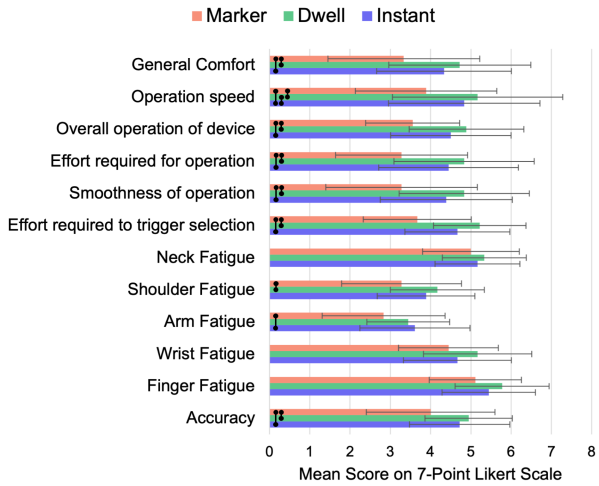



Fig. 8. Device Assessment Questionnaire scores by Selection Activation. Error bars show ± *SD*. Vertical bars show sig. diff. of $p < .001$.

$R^2 = 0.76$)). This yielded a strong positive correlation; regardless of selection activation or task difficulty, as tracking-loss count increased, so did tracking-loss duration.

*5) User Feedback:* After completing each condition, participants filled out the ISO Independent Assessment Questionnaire [17]. All Likert-scale questions had a 7-point range and higher scores are favourable. As seen in Fig. 8, Dwell on average scored highest in all categories except arm fatigue, While Marker on average scored lowest, 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. Instant and Dwell were by far the most preferred being ranked first by 9 and 7 participants and ranked second by 8 and 9 participants respectively. This aligns with the questionnaire results and participants' objective performance. Instant was overall most preferred.

Participants were 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 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.

## V. Experiment 2

Our second experiment used the same methodology as the first to compare Ray-cast, Head-gaze, and Virtual Hand selection techniques, all using dwell. Our main objectives were to evaluate the performance and UX of the Cardboard Controller with Virtual Hand selection of in-reach targets, and compare the controller's performance to Head-gaze, the most common MVR selection technique. We facilitated comparison between the two experiments by having corresponding conditions.

### A. Participants

We recruited 20 participants by poster and through university communications including 9 women and 11 men aged 18 and over (*M:* 25.15, $SD = 7.9$). The majority of participants tried VR at least once prior to the study, while three had no experience and four considered themselves expert users. Nine participants had prior experience with cardboard VR. All participants had normal or corrected vision except two who opted to not wear their glasses for comfort. All participants had prior experience using a spatial input device.

### B. Apparatus

*1) Hardware:* We used the same hardware in this experiment as the first, except the central marker was shrunk to match the size of the other two markers and its images were updated to improve tracking.

*2) Software:* We modified the software for this experiment, removing the red spheres representing the controller's tracked markers to more accurately simulate a real-world application. We set the selection tasks at two distances, in-reach (IR) at 50 cm, and out-of-reach (OR) at 3 m in front of the participant. The OR distance was the same distance as that used in the first experiment. Since selection activation was not under investigation in this study, all selections used a 300 ms dwell time. Thus, we eliminated the timed delay between selections used in the previous experiment.

OR targets were presented solely with the Ray-cast or Head-gaze techniques, as both support remote selection. We implemented Ray-cast the same way as in E1. Head-gaze selection used a ring cursor that was always rendered at the depth users were currently focused on. Since the cursor depth matched the target depth, the likelihood of participants experiencing double vision was low. The ring cursor appeared as a dot until it was pointed at a selectable target. It would then expand to a ring over 300 ms, at which time, selection occurred. The ring cursor was *not* subject to perspective scaling.

To provide selection feedback, regardless of technique, the target colour gradually shifted from purple to orange over 300 ms (see Fig. 9). Then, a sound effect played when a selection was

Fig. 9. Active target changing colour as gaze cursor hovers over it; colour shifts from purple to orange over 300 ms.

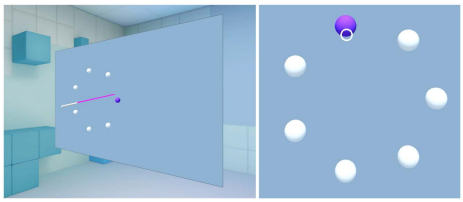


Fig. 10. Background panel directly behind IR targets in Ray and Gaze conditions. Selection ray and head-gaze ring cursor rendered on panel surface.

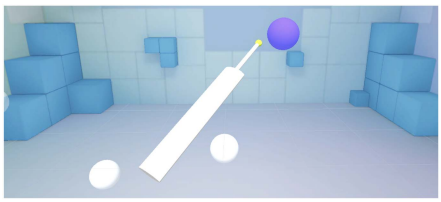


Fig. 11. Virtual controller in E2. Controller has an elliptical cylinder shape. For VH condition, selection point is highlighted in yellow.

triggered. The pitch of the sound effect increased with each subsequent selection in a ring of targets, resetting with the start of each new ring to the base pitch; this crescendo indicated to participants that they were making progress, to help make the task feel less menial.

All selection techniques were used to select IR targets. With the Ray-cast and Head-gaze conditions, a visible panel was placed behind the IR targets. The ray stopped being rendered once hitting the panel, and the Head-gaze cursor was rendered on the surface of the panel (see Fig. 10). The panel was added so there was not a stark difference between pointing the ray-cast or gaze cursor at the back wall of the VE versus an IR target. The panel is necessary for the Head-gaze condition since the depth of the back wall is significantly different from an IR target and caused double vision to occur when the user had to choose to focus one or the other.

The controller 3D model was changed from E1 to a thin rectangular block, so the mesh did not occlude the participants' view when using the Virtual Hand (VH) technique. We also added a thin rod and yellow ball to the end of the virtual controller (see Fig. 11). During the VH condition, the selection "hotspot" was at the tip of the ball and had the same width and height as the lines displayed with Ray-cast and Head-gaze. Thus, all techniques had the same size selection points. The active target was also made slightly transparent with the VH technique to help users judge the depth of the "cursor" and accurately place it in the centre of the target.

TABLE II
STUDY CONDITIONS

| | 50 cm | 300 cm |
|---|---|---|
| Head-gaze | Gaze-IR | Gaze-OR |
| Ray-cast | Ray-IR | Ray-OR |
| Virtual Hand | VH | |

Selection technique × depth.

TABLE III
AMPLITUDES AND TARGET WIDTHS IN CM UNDER EACH DEPTH

| | 50 cm | 300 cm |
|---|---|---|
| Amplitude (cm) | 22, 32, 40 | 110, 160, 200 |
| Width (cm) | 1.6, 3.2 | 8, 16 |

### C. Procedure

This procedure was similar to E1. We first described the procedure and how to use the controller. Participants then filled out a demographic questionnaire. Participants sat in a swivel chair in a well-lit room. Participants completed 12 practice trials that included all five conditions. Once finished, participants completed the five study conditions in counterbalanced order. After a participant finished all trials for a given selection technique, they filled out the ISO Independent Assessment Questionnaire [17] on Google Forms. Participants could take breaks between rounds. After all conditions, participants completed a questionnaire rating selection techniques in different scenarios, and soliciting subjective written feedback on the controller with each technique.

### D. Design

The experiment employed a within-subjects design with the following independent variables and levels:

- *Selection Technique:* Gaze-OR, Gaze-IR, Ray-OR, Ray-IR, VH
- *ID:* 2.98, 3.46, 3.76, 3.88, 4.39, 4.7
- *Selections:* 7
- *Repetitions:* 2

The five conditions are seen in Table II, organized by selection technique and target depth.

The six *ID*s were repeated twice with each condition. The amplitudes and target widths used in each depth are seen in Table III. *A* and *W* were the same as in the first experiment for the OR depth and were scaled down for IR target selections to match the same visual angle at both depths. Like with E1, we refer to the smaller widths as small targets and the larger widths as large targets.

We randomized *ID* order, and counterbalanced condition order according to a Latin square. In total, each participant completed 6 *ID*s × 2 repetitions × 5 selection techniques × 7 selections each, for 420 recorded selections. We recorded the same six dependent variables as in the first experiment: selection time (ms), target re-entries, selection distance from target centre (%), tracking-loss duration (ms), tracking-loss count, and throughput (bps).

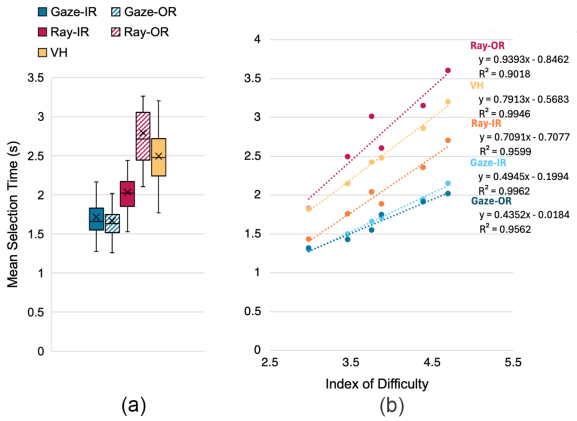



Fig. 12. (a) Mean selection time by condition. Sig. diff. of $p <. 001$ between all pairs. (b) Mean selection time by condition.

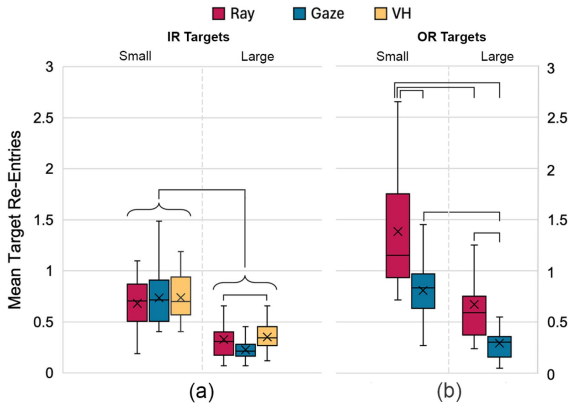



Fig. 13. Mean target re-entries by selection technique and target width. (a) in-reach conditions (b) out-of reach conditions. Sig. diff. of $p <. 001$ indicated.

TABLE IV
AVERAGE PERCENTAGE AWAY FROM TARGET CENTRE

| VH | Gaze-IR | Gaze-OR | Ray-IR | Ray-OR |
|---|---|---|---|---|
| 30.8% | 25.6% | 24.6% | 24.9% | 25.7% |

### E. Results

We recorded a total of 8400 selections. We removed 15 selections that timed out at 30 s and an additional 144 selections identified as outliers. We considered an outlier to be $\pm$ 3 *SD* from the mean for each independent variable combination. The 159 data points removed before analysis corresponded to 1.89% of the data. We used a Lilliefors normality test of each recorded dependent variable; all rejected the null hypothesis, thus we analyzed all metrics using a Friedman test with effect sizes represented with Kendall's *W*. All post-hoc analyses used Conover's F ($\alpha =. 05$). We performed power analyses using the same approach as in E1.

*1) Selection Time:* Mean selection times by condition are seen in Fig. 12. The Dwell condition from E1 that is identical to Ray-OR was added for means of comparison, but is not included in the following analysis. A Friedman test found a significant main effect for condition ($\chi^2 = 70.88$, $p <. 001$, $df = 4$, $W =. 1$, *pow.* $=. 60$). Posthoc tests revealed all pairs as significantly different. We modelled the relationship between selection time and *ID* (see Fig. 12) and found all conditions to be highly linear with all $R^2$ values surpassing 0.9. Gaze-IR and VH had the best fit with $R^2$ values above 0.99.

*2) Accuracy:* Accuracy was determined by selection distance, target re-entries, and time-outs that occurred after 30 s from the beginning of the selection inclusive of the controller losing tracking. In total, 15 selections were time-outs, equating to 0.18% of selections. An additional 21 data points were removed for the target re-entry analysis only. The removed data was extreme values about 3 *SD* of the mean. Mean target re-entries by condition and target width are seen in Fig. 13.

To explore interaction effects between target width and selection technique on re-entries, we performed an Aligned Ranked Transform ANOVA on the IR and OR data. In both depths, no significant interaction effects were found. For IR selections, we found a main effect of target width on target re-entries ($F_{(1,19)} = 322.69$, $p <. 001$, $\eta_p^2 =. 94$, *pow.* $=. 98$); smaller targets yielded significantly more re-entries. Significant pairs identified by posthoc tests are shown in Fig. 13.

For OR selections, we found a main effect of target width ($F_{(1,19)} = 235.47, p <. 001, \eta_p^2 =. 93$, *pow.* $=. 97$) and selection technique ($F_{(1,19)} = 38.39$, $p <. 001$, $\eta_p^2 =. 67$, *pow.* $=. 69$) on target re-entries. It was more difficult for participants to make accurate selections of far-away targets with the controller. We also observed that participants' gaze direction (the direction the camera is pointing) was more often pointed at the controller during IR tasks. Therefore, the camera direction relative to the controller's position may have an impact on tracking jitter and selection accuracy.

There was a significant main effect for IR selection technique on selection distance ($\chi^2 = 30.0$, $p <. 001$, $df = 2$, w $=. 75$, *pow.* $=. 70$). Posthoc tests revealed VH yielded significantly less accurate selections than Gaze-IR or Ray-IR potentially due to the increased control users had over the selection point location. When comparing OR selection techniques, no significant differences were found ($\chi^2 = 3.20$, $p =. 074$, $df = 1$ w $=. 16$, *pow.* $=. 15$). These results indicate Ray-cast and Head-gaze selection yield comparable accuracy under similar conditions.

Table IV shows on average how close selections were to the target centre under each condition. The table shows the percentage the average selection is from the centre, with 0% being perfectly on centre and 100% being on the target's edge.

*3) Throughput:* Since a primary benefit of using throughput is its consistency, we compared the *TP* scores for a condition that appeared in both experiments (Ray-OR and dwell activation). Results of two Mann-Whitney U tests using a two-tailed hypothesis found no significant differences between euclidean *TP* ($z =. 45$, $p = 0.65$, $r =. 074$, *pow.* $=. 05$) or projection-based *TP* ($z = 0.13$, $p = 0.90$, $r = .021$, *pow.* $=. 09$), indicating consistency across experiments.

Fig. 14 shows the average scores across conditions. Two Friedman tests showed the effect of selection technique on both euclidean-distance ($\chi^2 = 67.92$, $p <. 001$, $df = 4$, w $=. 85$, *pow.* $=. 59$) and project-based *TP* ($\chi^2 = 67.24$, $p <. 001$, $df = 4$, $W =. 84$, *pow.* $=. 59$) were significant. Posthoc tests revealed significant differences between all pairs except Ray-OR and VH for both *TP* calculations.

*4) Tracking Performance:* To evaluate tracking performance, we recorded the elapsed time the controller lost tracking

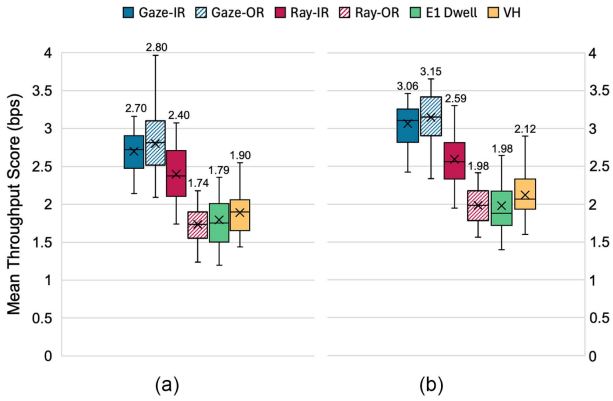



Fig. 14. Mean *TP* scores by condition; (a) projection-based (b) euclidean-distance. Sig. diff. of $p <.001$ between all pairs except Ray-OR and VH.

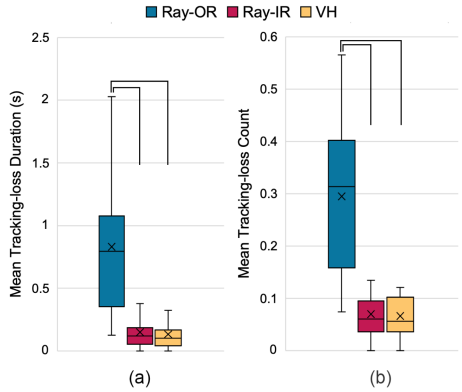



Fig. 15. Mean tracking-loss duration (a) and tracking-loss count (b) by selection technique. Sig. diff. of $p <.001$ indicated.

and the number of times tracking was lost during a selection. We performed Friedman tests on tracking-loss time ($\chi^2 = 27.10$, $p <.001$, $df = 2$, w $=.68$, *pow.* $=.60$) and tracking-loss count ($\chi^2 = 22.56$, $p <.001$, $df = 2$, $W =.56$, *pow.* $=.61$). For both metrics, posthoc tests revealed it was significantly more difficult to maintain tracking given OR targets, while no significant differences were found between IR-target conditions (see Fig. 15). Similar to re-entries, this difference in tracking reliability likely stems from IR tasks forcing users to keep the controller in front of the camera.

*5) User Feedback:* After completing all conditions with a given selection technique, participants filled out the ISO Independent Assessment Questionnaire. Head-gaze scored highest in all categories except neck fatigue (see Fig. 16). VH scored above Ray-cast in all categories except arm, shoulder, and neck fatigue. No significant differences between VH and Ray-cast scores were found except for shoulder fatigue.

At the end of the experiment, we asked participants their preference of selection technique for IR and OR selection tasks. Head-gaze was favoured overall while VH was favoured over Ray-cast for IR tasks (see Table V). Participants supplied written responses describing their choices. We reviewed all responses and identified reoccurring themes.

Participants who favoured Head-gaze for OR selections often stated it felt more accurate, faster, and easier to operate. Meanwhile, five of 15 participants felt Ray-cast was more difficult to use because the controller would lose tracking. These findings align with the mean questionnaire scores. Most participants who preferred Ray-cast over Head-gaze felt it was less fatiguing to operate and more intuitive to control as handheld controllers are more familiar to them. Two participants stated selecting with the controller felt more purposeful; they liked pointing at specific locations with the ray-cast opposed to using head-gaze and possibly finding targets by "chance".

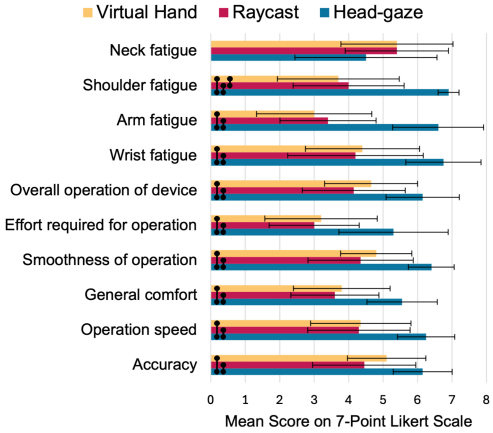



Fig. 16. Mean scores of ISO Questionnaire by selection technique. Error bars show ± *SD*. Vertical bars show sig. diff. of $p <.001$.

TABLE V
SELECTION TECHNIQUE PREFERENCE BY DEPTH AND OVERALL

| | Head-gaze | Ray-cast | Virtual Hand |
|---|---|---|---|
| IR | 9 | 3 | 8 |
| OR | 15 | 5 | |
| 1st | 14 | 3 | 3 |
| 2nd | 3 | 11 | 6 |
| 3rd | 3 | 6 | 11 |

In regard to IR selection tasks, all participants preferred a given technique because they perceived it as the easiest to operate and most accurate. Three participants preferred VH because it felt the most "interactive"; they liked physically reaching out to targets and seeing them respond to their touch.

We asked participants to rate each selection technique overall and Head-gaze was most preferred, followed by Ray-cast (see Table V). We also asked participants for their opinion on operating the controller. The most common response, stated by six participants, were remarks on the tracking limitations when using Ray-cast for OR selections. One participant stated it was difficult to point the ray near the bottom of their FOV, while a second participant noted speed and angle limitations; if the controller was moved too quickly or held at an extreme angle, it would lose tracking. Three participants stated they preferred Ray-cast *when* the tracking was reliable. Of those, one participant claimed if tracking is improved, they would change their preference from Head-gaze to the controller.

## VI. DISCUSSION

*1) Selection:* Our first experiment explored potential selection activation methods for the Cardboard Controller, comparing manual (Marker) and automatic (Instant, Dwell) input. Instant and Dwell yielded faster selection times on average (see Fig. 4),

consistent with past findings [13]. Although instant produced the fastest selection times, it is not a realistic technique for real-world applications; it yields unintended selections [18]. With Marker, most participants occluded the selection marker with a cautious hand movement, likely exaggerating selection times. We attribute these slower movements to the unfamiliarity of Marker and the controller's less robust materials. With more practice, we speculate selection times with manual input may improve.

There was a visible negative correlation between selection time and tracking loss (see Figs. 4 and 7). While Marker had the longest selection times on average, it also begot the highest tracking-loss times and count. With Marker, participants had to keep two objects tracked simultaneously within a small FOV making unintended occlusions likely. This adds complexity to the otherwise simple selection task and may account for the increase in tracking loss and selection times.

Based on these results, Marker is not an effective method for confirming selections with the Cardboard Controller, addressing R1, and manual input must be re-thought (see Future Work). Nonetheless, selection times using the Cardboard Controller with dwell activation ($M = 2.3$ s) show promise as they are comparable to past evaluations of similar 3D selection techniques [32], [41]. These findings are reinforced by our second experiment where mean selection times ranged from 2-2.79 s. Pham et al. [32] evaluated an HTC Vive controller with space bar input and had average selection times slightly over 1 s. Meanwhile, Yu et al. [41] evaluated pointing and virtual hand selection with button and grasp gesture input with average selection times of 2.38 s and 2.33 s respectively.

The second experiment focused on the controller's efficacy for supporting ray-casting and virtual hands compared to head-gaze selection as a baseline. Overall, head-gaze outperformed the Cardboard Controller in selection time and *TP* scores at both target depths. When targets were in-reach, the ray was significantly faster than the virtual hand. This aligns with past findings of pointing outperforming other selection metaphors [1]. Despite its worse performance, 8 of 11 participants favoured the virtual hand technique over ray-casting for IR tasks (See Table V). Participants noted the virtual hand technique felt more interactive and purposeful; they enjoyed reaching toward targets over passively pointing.

*2) Accuracy:* Notably, Marker activation produced selections closer to target centre than Dwell or Instant. This was not surprising as manual input provides more control over when and where a selection is confirmed. The bimanual operation of Marker also helps mitigate the Heisenberg effect [3]. While the second experiment used dwell time for input with all selection techniques, we still found significant differences in selection accuracy. For near selections, the virtual hand was least accurate; participants had to estimate the depth of the cursor and target, posing an additional challenge. In contrast, the ray-based techniques (Ray-IR and Gaze-IR) computed the closest point on the ray to target centre automatically. When comparing the identical conditions between experiments, accuracy was very similar (26.34% and 25.5% away from target centre). This further confirms homogeny across experiments.

In both experiments, tracking loss and jitter affected target re-entries. The controller's tracking was unstable when held at the edges of the FOV or at less natural orientations. This effect was most noticeable in the Ray-OR condition and it aligns with prior knowledge of the shortcomings of ray-based techniques; they can be difficult to orient precisely to select far-away or small objects [33]. In the second experiment, Ray-IR and VH were less affected by tracking issues as the camera tended to stay directed at the controller throughout the task.

*3) Throughput:* Throughput scores remained consistent across both experiments (See Fig. 12). In the Ray-OR condition, the Cardboard Controller yielded average scores from 1.74 to 1.98 bps. For IR tasks, scores notably improved to 2.59 and 2.4 bps depending on the calculation method. Ray-OR and VH *TP* scores were comparable and lower than Ray-IR, suggesting the ray-cast technique yields better performance than the virtual hand technique under similar conditions.

As detailed above, throughput provides a reliable means of comparison across studies [36]. Past studies evaluating similar ray-based techniques to the Cardboard Controller reported *TP* scores ranging from ∼1.5-2.5 bps [35], [37], [39]. For example, an HTC Vive controller evaluated in a 3D selection task yielded *TP* of 1.39 bps [39].

The Cardboard Controller is in the range of 1.5-2.5 bps with dwell-based selection activation, and exceeded it in experiment 2's Ray-IR condition depending on the calculation method. The results answer R2 and R3 and we take this as evidence that under the right circumstances, the Cardboard Controller offers sufficient performance as an input device for MVR and can effectively support both pointing and virtual hand interaction.

*4) Best Technique for Cardboard Controller:* Overall, ray-casting outperformed the virtual hand technique in selection, accuracy, and throughput for IR tasks, but virtual hand was far preferred by participants. Our results suggest the Cardboard Controller is most effective for IR tasks. For OR tasks, Gaze-OR outperformed Ray-OR and was most preferred by participants. To maintain tracking during OR tasks, the controller needed to be held in a higher, less natural position, amplifying the known '*gorilla arm effect*' (arm fatigue caused by mid-air interactions) [2]. To answer R4, the Cardboard Controller can effectively support both techniques, but virtual hand interaction may be better suited as the tracking is more stable for IR tasks and the technique was preferred by participants.

## VII. Future Work

We plan to evaluate new manual input methods to replace the Marker method. A one-handed solution is ideal, and we have begun early work investigating acoustic sensing to trigger selection activation via sound [11]. We also plan to evaluate the Cardboard Controller's performance in object manipulation tasks; we speculate object manipulation could be performed via a click-and-hold action. We plan to improve the usability of the controller for OR tasks by enhancing the quality of the controller's tracking. This may involve adding a gyroscope and updating marker layout. Lastly, we plan to design a more ergonomic handle for the controller akin to a wand controller.

## VIII. Study Limitations

The two-second delay between selections in E1 (see Section IV-B2) 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. Similarly, because each subsequent target was not made active instantly, this introduced a delayed reaction time to the beginning of each recorded selection time. Although both experiments lacked a baseline condition, technical limitations prevent a reasonable comparison without introducing numerous confounding variables (e.g., comparing a Meta Quest + controller to the Google Cardboard + Cardboard controller). Finally, both experiments employed a modern (at time of writing) smartphone; an older device may yield worse performance and user experience.

## IX. Guidelines for Designing Mobile VR Input Devices and Interactions

The following section outlines best practices for developing low-cost input devices or interaction methods for MVR. We compiled this list based on the findings of past research and our own. It is our hope to contribute to further developments towards high-complexity, low-cost interactions for mobile VR.

*1. Take Advantage of the Materials:* While a cardboard viewer is low fidelity, the smartphone encased inside is not and can be leveraged in creative ways to afford more complex interactions. For example, Chen et al. [5] and Yan et al. [40] both used smartphone functions in ways they were not originally intended for. Chen et al. used the microphone while Yan et al. used the phone's motion sensors to detect tap and swipe gestures on the cardboard surfaces. Other parts of the phone could also be employed such as the front camera [7], depth camera on newer models, touch screen, or other sensors (e.g., accelerator or compass altitude sensor for travel).

*2. Creativity is Key* . Designing for affordability and convenience severely limits possible materials and features available. One cannot fully rely on the technology and must think out of the box to conjure creative solutions to circumvent the longstanding limitations of MVR. Past work [5], [23], [40], and the current work, have utilized the smartphone camera or cardboard viewer creatively to develop new input methods.

*3. Diversity in Tracking:* Input device tracking performance can be improved by detecting controller pose from multiple and/or redundant sources when possible. Optical tracking is most commonly used to support MVR interactions [6], [9], [24], [27]; however, it has known issues caused by poor lighting conditions or a limited FOV [9], [24]. Multiple tracked points and/or trackers provide backups in the case of tracking failure. While the three markers on our Cardboard Controller afforded more consistent tracking than a single one used in initial prototyping [10], additional pose data from a second source (e.g., a gyroscope) may further improve tracking consistency. Mohr et al. [27] similarly tracked a handheld device optically while simultaneously tracking the HMD using the device; the pose information was synchronized.

*4. Role of Environment:* If a tracking system is used, we suggest testing the device performance in different environmental conditions and planning for environmental inconsistencies. Depending on the tracking technology, lighting, sound, or other elements can disturb the system. The Cardboard Controller uses optical tracking and optimally performs when used in bright lighting, but the ideal MVR input device should work in suboptimal conditions as its intended for personal use.

*5. Consider Processing Power of Smartphone:* A well-known limitation of MVR is the limited processing power of the phone [6], [7], [29]. When evaluating a new interaction method for MVR, it is important to consider the smartphone's limitations in the evaluation [7], as well as perform the evaluation on the intended device. For example, if hand tracking is employed using only the smartphone, the computational complexity required should be considered in the evaluation. Some past research presenting novel interaction methods for MVR failed to test their solutions on an MVR headset [12] or did not evaluate its computational complexity when relevant [29] making their findings less reliable.

## X. Conclusion

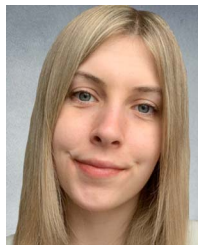
We present the design and evaluation of the Cardboard Controller — a costless 6DOF tracked controller for MVR — and design guidelines for future developments in MVR prototyping. We evaluated selection performance and user experience of the controller over two user studies. The first validated the controller as a 3D input device with throughput scores comparable to commercial VR controllers evaluated under the same methodology [39]. The second experiment demonstrated the controller's efficacy to support both pointing and virtual hand selection, especially for in-reach objects.

The Cardboard Controller's highly accessible materials and simple assembly gives great potential for consumer use and enables researchers to ship the materials to participants to be assembled remotely. This would reduce geographic limitations for participant recruitment and reach broader more representative participant populations that have otherwise historically lacked diversity [30]. The Cardboard Controller increases the complexity and variety of interactions supported in MVR previously only possible with expensive devices and is a step towards the democratization of VR.

## References

[1] F. Argelaguet and C. Andujar, "A survey of 3D object selection techniques for virtual environments," *Comput. Graph.*, vol. 37, pp. 121–136, 2013.
[2] S. Boring, M. Jurmu, and A. Butz, "Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays," in *Proc. Annu. Conf. Aust. HCI Special Int. Group*, pp. 161–168, 2009.
[3] D. Bowman, C. Wingrave, J. Campbell, and V. Ly, "Using Pinch Gloves for both Natural and Abstract Interaction Techniques in Virtual Environments," Comp. Sci., Virginia Tech, VA, USA, Tech. Rep. TR-01-23, 2001.
[4] D. Bowman and L. Hodges, "Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments," *J. Vis. Lang. Comput.*, vol. 10, pp. 37–53, 1999.
[5] T. Chen, L. Xu, X. Xu, and K. Zhu, "GestOnHMD: Enabling gesture-based interaction on low-cost VR head-mounted display," *IEEE Trans. Visualization Comput. Graphics*, vol. 27, pp. 2597–2607, 2021.
[6] M. Castro, A. J. Xavier, P. Rosa, and J. Oliveira, "Interaction by hand-tracking in a VR environment," in *Proc. IEEE Int. Conf. Ind. Technol.*, 2021, pp. 895–900.
[7] P. Drakopoulos, M. George-Koulieris, and K. Mania, "Eye tracking interaction on unmodified mobile VR headsets using the selfie camera," *ACM Trans. Appl. Percep.*, vol. 18, pp. 1–20, 2021.
[8] W. Greenwald, "The best VR headsets for 2024," 2024. [Online]. Available: https://www.pcmag.com/picks/the-best-vr-headsets
[9] K. Grinyer and R. Teather, "Low-fi VR controller: Improved MVR interaction via camera-based tracking," in *Proc. ACM Symp. Spatial User Interact.*, 2023, pp. 1–12.

[10] K. Grinyer and R. Teather, "Rapid prototyping of low-fidelity VR hardware: Lessons learned," in *Proc. 2024 IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2024, pp. 415–418.
[11] K. Grinyer and R. Teather, "ClickSense: A low-cost tangible active user input method using passive acoustic sensing for mobile VR," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2025, pp. 1–7.
[12] J. Gugenheimer, D. Dobbelstein, C. Winkler, G. Haas, and E. Rukzio, "FaceTouch: Enabling touch interaction in display fixed UIs for mobile VR," in *Proc. 29th Annu. Symp. User Interface Softw. Technol.*, 2016, pp. 49–60.
[13] J. Hansen, V. Rajanna, I. MacKenzie, and P. Bækgaard, "A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display," in *Proc. Workshop Commun. Gaze Interact.*, 2018, pp. 1–5.
[14] T. Hirzle, J. Rixen, J. Gugenheimer, and E. Rukzio, "WatchVR: Exploring the usage of a smartwatch for interaction in mobile VR," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2018, pp. 1–6.
[15] D. Hoffman, T. Novak, A. Schlosser, and B. Compaine, *The Digital Divide. Facing a Crisis or Creating a Myth*. Cambridge, MA, USA: MIT Press, 2001.
[16] D. Hufnal, E. Osborne, T. Johnson, and C. Yildirim, "The impact of controller type on video game UX in VR," in *Proc. IEEE Games, Entertainment, Media Conf.*, 2019. pp. 1–9.
[17] *Ergonomics of human-system interaction — Part 411: Evaluation methods for the design of physical input devices*, ISO/TS 9241-411:2012, ISO, 2012. [Online]. Available: https://www.iso.org/standard/54106.html
[18] R. Jacob, "What you look at is what you get: Eye movement-based interaction techniques," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 1990, pp. 11–18.
[19] D. Kharlamov, B. Woodard, L. Tahai, and K. Pietroszek, "TickTockRay: Smartwatch-based 3D pointing for smartphone-based VR," in *Proc. ACM Conf. VR Softw. Tech.*, 2016, pp. 365–366.
[20] Y. Kim, R. Ilsun, M. Hwan, and Y. A. Systematic, "Review of a VR system from the perspective of UX," *Int. J. Hum.–Comput. Interact.*, vol. 36, 2020, pp. 93–910.
[21] S. Kyian and R. Teather, "Selection performance using a smartphone in VR with redirected input," in *Proc. 2021 ACM Symp. Spatial User Interact.*, 2021, pp. 1–12.
[22] E. Lehmann, and H. D'Abrera *Nonparametrics: Statistical Methods Based on Ranks*. Berlin, Germany: Springer, 2006.
[23] R. Li, V. Chen, G. Reyes, and T. Starner, "ScratchVR: Low-cost, calibration-free sensing for tactile input on mobile VR enclosures," in *Proc. ACM Int. Symp. Wearable Comput.*, 2018, pp. 176–179.
[24] S. Luo, R. Teather, and V. McArthur, "Camera-based selection with cardboard head-mounted displays," in *Proc. Int. Conf. Hum.-Comput. Interact.*, 2020, pp. 383–402.
[25] I. MacKenzie, *The Wiley Handbook of Human Computer Interaction*. Hoboken, NJ, USA: Wiley, 2018.
[26] R. McGloin, K. Farrar, and M. Krcmar, "The impact of controller naturalness on spatial presence, gamer enjoyment, and perceived realism in a tennis simulation video game," *Presence: Teleoperators Virtual Environ.*, vol. 20, pp. 309–324, 2011.
[27] P. Mohr, M. Tatzgern, T. Langlotz, A. Lang, D. Schmalstieg, and D. Kalkofen, "TrackCap: Enabling smartphones for 3D interaction on mobile head-mounted displays," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–11.
[28] M. Cardboard, "POP! CARDBOARD 3.0," 2024. [Online]. Available: https://mrcardboard.eu/
[29] K. Park and J. Lee, "New design and comparative analysis of smartwatch metaphor-based hand gestures for 3D navigation in mobile VR," *Multimedia Tools Appl.*, vol. 78, pp. 6211–6231, 2019.
[30] T. Peck, K. McMullen, and J. Quarles, "DiVRsify: Break the cycle and develop VR for everyone," *IEEE Comput. Graph. Appl.*, vol. 41, no. 6, pp. 133–142, Nov./Dec. 2021.
[31] Pew Research Centre, "Mobile fact sheet," 2024. [Online]. Available: pewresearch.org/internet/fact-sheet/mobile/
[32] D. Pham and W. Stuerzlinger, "Is the pen mightier than the controller? a comparison of input devices for selection in virtual and augmented reality," in *Proc. 25th ACM Symp. Virtual Reality Softw. Technol.*, 2019, pp. 1–11.
[33] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst, "Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques," *Comput. Graph. Forum*, vol. 17, pp. 41–52, 1998.
[34] W. Powell, V. Powell, P. Brown, M. Cook, and J. Uddin, "Getting around in Google cardboard–exploring navigation preferences with low-cost mobile VR," in *Proc. IEEE Workshop Everyday VR*, 2016, pp. 5–8.
[35] A. Ramcharitar and R. Teather, "EZCursorVR: 2D selection with VR head-mounted displays," in *Proc. Graph. Interface Conf.*, 2018, pp. 123–130.
[36] R. Soukoreff and I. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI," *Int J. Hum.-Comput. Stud.*, vol. 61, pp. 751–789, 2004.
[37] R. Teather and W. Stuerzlinger, "Pointing at 3D target projections with one-eyed and stereo cursors," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2013, pp. 159–168.
[38] M. Terenti, C. Pamparău, and R. Vatavu, "The user experience of distal arm-level vibrotactile feedback for interactions with virtual versus physical displays," *Virtual Reality*, vol. 28, 2024, Art. no. 84.
[39] W. -J. Hou and X.-L. Chen, "Comparison of eye-based and controller-based selection in VR," *Int. J. Hum.–Comput. Interact.*, vol. 37, pp. 484–495, 2021.
[40] X. Yan, C. Fu, P. Mohan, and W. Goh, "CardboardSense: Interacting with DIY cardboard VR headset by tapping," in *Proc. ACM Conf. Designing Interactive Syst.*, 2016, pp. 229–233.
[41] D. Yu, H. Liang, F. Lu, V. Nanjappan, K. Papangelis, and W. Wang, "Target selection in head-mounted display VR environments," *J. Universal Comput. Sci.*, vol. 24, pp. 1217–1243, 2018.



**Kristen Grinyer** is currently working toward the PhD degree with Carleton University, in Ottawa, Canada. Her research interests include interactions and human performance in VR/XR environments and improving the accessibility of XR technologies.

**Robert J. Teather** is an associate professor in and director of the School of Information Technology with Carleton University. He is co-director of the MARVEL research group, and studies on human performance in mixed, augmented, and virtual reality environments.