Enabling Button-based Interaction in Mobile Virtual Reality Using Acoustic Sensing to Support 3D Pointing Techniques

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

Robert J. Teather
rob.teather@monash.edu
Monash University
Melbourne, Victoria, Australia

Abstract

Mobile virtual reality (MVR) facilitates financially accessible VR by incorporating a smartphone into a head-mounted viewer but lacks effective input mechanisms to trigger events. We propose a low-cost unpowered button detected via acoustic sensing and integrated into low-fi pointing devices to support interaction in MVR. The smartphone's built-in microphone passively listens to environmental audio to identify a unique click sound issued by the button. We compared button performance to 300 ms dwell time with both head-gaze pointing and an optically tracked cardboard controller. The button offered pointing throughput similar to ray-based selection techniques reported in prior studies at 2.22 bps with head-gaze, and 1.81 bps with the controller. Overall, the button accuracy rate was 92.5% with 7.85% of clicks registered as false negatives. Our results suggest the button is a viable input mechanism for MVR that can support 3D pointing techniques.

CCS Concepts

• **Human-centered computing → Virtual reality**; *Pointing devices*; **Sound-based input / output**; *Pointing*.

Keywords

Virtual Reality, Sound-based Input, Pointing Devices

ACM Reference Format:
Kristen Grinyer and Robert J. Teather. 2026. Enabling Button-based Interaction in Mobile Virtual Reality Using Acoustic Sensing to Support 3D Pointing Techniques. In *Proceedings of Graphics Interface (GI'26)*. ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Extended reality (XR) head-mounted displays (HMD), including virtual reality (VR), mixed reality, and augmented reality, have potential as future general purpose personal computing devices [39]. However, device cost remains a prominent adoption barrier [26] to people of lower socioeconomic statuses. This widens the digital divide—the disparity in access, usage behaviour, acquired skills, and ability to benefit from technology across socioeconomic groups [57, 58].

Mobile VR (MVR) offers a potential solution. MVR HMDs use a smartphone as a low-cost VR device. Since smartphones are already ubiquitous across income/education levels [51], MVR offers immersive experience to virtually anyone. Plastic and cardboard MVR viewers cost between $10–60 USD, far less than $500-3500 USD for high-end devices. Virtually all modern HMDs include a 6-degrees-of-freedom (DOF) tracked wand controller to support interaction (e.g., via virtual hands and pointing). However, MVR supports only 3DOF head tracking and primitive input such as single click, double click, or dwell-based pointing [14]. This "interaction gap" limits MVR's potential to bridge the digital divide [20].

Some plastic MVR devices include a non-tracked Bluetooth controller, but their availability and price are inconsistent and region dependent. Plastic viewers range in price from $25–65 USD or upwards of $70 USD. Zappar's mixed reality viewer[1] offers two 6DOF tracked controllers for $70 USD, but only supports IOS devices and is only available in the US, EU and UK. We focus exclusively on cardboard-based VR, as the most accessible MVR due to low cost, global availability, and support of both IOS and Android devices.

Without tracked controllers and/or hand-tracking, MVR cannot support common VR interaction techniques such as virtual hands or ray-based pointing [3, 35, 41], and thus cannot provide experiences comparable to high-end devices. Past research has focused on improved interaction for MVR, such as UI interaction [21, 24, 36, 36, 67, 67], eye tracking [2, 13], or hand tracking [12, 41] through low-cost means. However, most prior work on 3 or 6DOF pointing used expensive external devices such as a smartwatch or second smartphone [25, 33, 34, 44, 52], or found the integrated Google Cardboard button unreliable [5, 41].

The cardboard controller [16, 18] provides 6DOF pose information and supports ray-based selection in MVR, but lacks an active selection activation mechanism. Selection activation is the critical final step of a selection task; after positioning the cursor/ray on a target, some mechanism is required to indicate task completion [6]. Active selection activation methods—those requiring user action—offer better control than time-based methods (e.g., dwell) [14] and improve selection accuracy [16, 22].

We present a low-cost, unpowered, universally compatible button detected via passive acoustic sensing: when pressed, the button emits a click noise the system detects. Our solution leverages the affordances of an everyday object (a jar lid) and repurposes it for input. The design is socioeconomically accessible and sustainable; it is battery-free, low-energy, and reduces e-waste. Our informal evaluation found the button offers a 97% and 95% recognition rate used in isolation and a standard selection task respectively (see Section 3). To evaluate the button for selection activation, we attached

GI'26, Waterloo, ON

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

[1] https://www.zappar.com/zapbox/

it to a cardboard controller and compared it to dwell using both controller- and head-based ray pointing. Head-gaze is the most common interaction technique for MVR, where the user selects targets by directing their head towards them [54]. Participants used both pointing techniques with both selection activation methods in a study conforming to the ISO-9241-411 standard [29].

The main contributions of our work include:

- A novel active user input mechanism for use with new and existing MVR interaction techniques.
- Further insight into the use of passive acoustic sensing for low-cost interaction in VR.
- A formal experiment showing evidence our button offers reasonable performance in selection tasks.

2 Related Work

We reviewed past research on selection techniques and selection activation methods in VR, as well as passive acoustic sensing.

2.1 Selection in VR

Selection in VR is a fundamental task that typically precedes other tasks, such as object manipulation or system control [3, 11, 35]. Selection technique effectiveness has a profound effect on user performance and experience [3]. Selection techniques are often classified by metaphor [3, 6, 53]; the most common metaphors are virtual hands and ray-based pointing [3, 35, 41]. Virtual hands present a hand or proxy controlled by a tracked handheld device or user's hands. Objects are selected by directly intersecting objects with the virtual proxy [6, 37]. Pointing techniques support out-of-reach selection by intersecting objects with a ray whose orientation and (usually) origin are controlled by the user [35]. While selection metaphors describe *how* a user specifies a target, most selection techniques require some action to confirm selection [8]. This selection activation step was identified early in VR research as a critical final step of a selection task [6].

2.2 Selection Activation Methods in Mobile VR

Selection activation commonly employs physical controls, sound or voice, gestures, or passive methods (i.e., dwell time) [6, 35]. Choosing an appropriate method is important, as it affects users' sense of control. This is challenging in MVR due to its fewer activation options [14, 16, 41, 54], including click, double click, and dwell.

Physical controls (e.g., buttons), built into the input device are reliable and provide tactile feedback [27]. However, in MVR, built-in magnetic or conductive buttons are unreliable [5, 41] and susceptible to the Heisenberg Effect [33] (unintended cursor movement at the instant of selection [7] due to the button press shifting the device). In contrast, auditory input and gestures support hands-free interaction. Mid-air hand gestures are more engaging and faster than handheld devices [41], but yield fatigue [4, 9, 67]. However, neither gestures nor sound input match the accuracy and reliability of physical controls due to optical registration errors [33, 41, 50] and environmental audio interference [23, 75], respectively. Camera-based gesture detection imprecision worsens in MVR due to the limited field of view (FOV) and processing power [4, 16, 17, 41]. Passive dwell time is common in MVR and requires the user to hover or point a ray at an object; selection triggers automatically after a time-out [14, 46]. Passive input offers less control than active input, increasing likelihood of selection errors [14, 16, 22].

Previous work compared 300 ms dwell to button input [14], a VR controller trackpad, and pinch gestures [46]. While dwell was slowest, it yielded fewer errors. Dwell selection times were comparable to clicks with the Google Cardboard button, yet participants preferred the button, despite fatigue [14], a well-known problem with mid-air interaction [4, 21, 41]. Our proposed button mitigates arm fatigue as it allows instantaneous selection activation, avoiding the "gorilla arm effect" caused by both the Google Cardboard button and dwell when paired with a handheld controller.

2.2.1 *Emerging Activation Methods.*

Past MVR research used extra hardware to support activation methods. Past work explored mounting a touchscreen to the front of the HMD to support tapping and sliding gestures[21, 36]. Other work used handheld devices such as smartwatches or smartphones for tap interaction [25, 34, 44] or to detect forearm rotation [33] or finger-pointing gestures [52]. These approaches require extra hardware and cost.

Other work employed optical tracking to support hand gestures for selection activation [12, 41]. De Castro et al. [12] offloaded gesture recognition processing to an external PC, but this solution was not evaluated. Previous work found both mid-air finger taps and the Google Cardboard button performed worse than a touchpad button [4, 41], and camera-based tracking unreliable for selection activation [16, 41], revealing an advantage of physical controls.

Low-cost "DIY" approaches have been explored, such as using the smartphone camera to detect a state change in tracked paper markers (e.g., occlusion) for selection activation [16]. This approach performed worse than dwell due to limited camera FOV size and tracking reliability. Other input approaches used low-cost materials and the smartphone's speaker and sensors. Past work detected touch events on the smartphone screen using conductive patterns placed on the sides and front of a Google Cardboard [14, 32]. The smartphone microphone [9] or accelerometer [71] have been used to detect sliding gestures on the sides of a Google Cardboard. Meanwhile, BlowClick [77] relied on users blowing into the microphone to trigger events; it had a higher learning curve than a controller button and caused fatigue. Our current work was inspired by previous research using microphone data for input [9, 77].

2.3 Sound Recognition

Sound-based input dates back to the 1950s, with the Zenith Space Command TV remote[2], which emitted ultrasonic frequencies from tuning forks struck by spring-loaded hammers upon pressing remote switches. A microphone in the TV detected the frequencies which were mapped to operations like changing channels. Although prone to environmental noise, it shows that low-fidelity clicking mechanisms, like our approach, can support simple user input.

Sound detection requires acoustically unique sounds [23]. The simple 1D data and filtering lend themselves well to MVR as higher computation yields lag, system crashes, and can overheat the device. However, it also limits the possible input options [60]. Passive acoustic sensing has been used to recognize sounds produced by tangible objects [62, 63] and sliding and tap gestures on surfaces

[2]https://www.theverge.com/23810061/zenith-space-command-remote-control-buttonof-the-month

[8, 9, 23, 24, 38, 40, 60, 74]. Scratch Input [23] detected unique acoustic signals produced from scratching a fingernail on a surface achieving an 89.5% accuracy rate. Scratch Input inspired several other projects [8, 9, 24, 38, 40, 60, 63, 70, 74], including our work. Section 3 details similar filtering strategies we used to isolate sound produced by the button to reduce false positives.

3 System Design

We used a custom cardboard controller made of a folded cardboard handle and three tracked cubic markers (see Figure 1b). A detailed system design and evaluation are reported in prior work [16, 18].

Given issues with camera-based tracking in MVR [4, 12, 16, 41], our selection activation method instead uses physical controls and sound. While optical tracking depends on smartphone camera quality, sound detection is consistently accurate across smartphone models [9]. Our system provides the tactile feedback and reliability of a physical button [27], and is both familiar and intuitive. It does not require either wired or wireless connectivity nor additional complex/expensive hardware. While Bluetooth controllers are an inexpensive option for button-based input, they introduce pairing, connectivity, and battery issues, which impacts their usability [1, 55, 69]. They also contribute to e-waste. Our button circumvents these issues as it is non-powered, requires no pairing step, and is universally compatible and reliably works with any smartphone. Moreover, the button design negates issues related to over-heating and weight, and does not rely on existing sensors of the smartphone.

3.1 Prototyping Process

When exploring potential sound sources, we used household materials to keep costs low and focused on generating a sound waveform that is 1) easily and consistently replicated by users, 2) unique and distinctive from speech and ambient noise, and 3) can be made with one hand. When testing objects, we found homemade shakers too inconsistent, dog clickers too shrill, but found the convex shape of a metal jar lid makes a distinct popping noise when pressed down on two points. Full details of our prototyping process are in prior work [19]. To support one-handed operation, we made a device using the cardboard controller [16, 18], jar lid, and keyring. See Figure 1b.

To identify the unique acoustic pattern produced by the button, we developed a Unity application that passively records audio using the smartphone's microphone and sends the audio amplitude pattern to an external spreadsheet. We used this data to define filters for isolating and detecting the click sound. To evaluate and refine these filters, we collected data in two phases to first define filter values for sound recognition then refine those values to improve detection accuracy. We used a Samsung Galaxy S23 Ultra, a Google Cardboard, and the button prototype (Figure 1a) for all tests.

To collect data, the microphone passively listened for and recorded audio sample data. Unity audio sample data is contained in an array of floats ranging from -1.0 to 1.0. While these values are arbitrary, they mimic the audio's amplitude pattern and indicate its volume. As such, we refer to the mean value of the sample data array as the volume of that sample. Like prior work [23], to reduce false positives and filter out environmental noise, we ignored sample data with an absolute mean value below 0.05. The microphone registered a new audio clip when the volume rose above 0.05, ending it once it fell below that threshold. Audio clips with durations less than 100 ms were ignored to further filter out ambient noise.

(a) (b)

Figure 1: (a) Button prototype made of cardboard, metal lid, and key ring; user presses down to elicit a click. (b) Cardboard controller with button apparatus.

We recorded 300 sound samples (clicks) in a quiet environment for both data collection phases. To record a sample, we pressed the prototype button in short intervals. For each sample, we recorded amplitude variation metrics known to be effective for sound recognition [23] including duration in milliseconds, peak volume value, and the timestamp of when the peak value occurred. Using the data from phase 1, we defined values for the three filters listed above. In phase 2, we used these filters and recorded whether the sound passed all filters or was a false positive or negative. Of the 300 recorded samples, 25 clicks (8.34%) were not registered as input. Based on these findings, we adjusted the filter values so 97.33% of the 300 samples correctly registered as input.

3.2 Final Design

We fastened the jar lid apparatus to the controller and the key ring was attached to an additional cardboard piece with two extra cardboard layers for strength. We cut a slit in the top of the controller and slotted the button apparatus in, aligning the key ring and lid properly. Figure 1b shows the controller used in the experiment.

4 Methodology

We conducted an experiment conforming to ISO 9241-411 [29] based on Fitts' law [48], which models the linear relationship between movement time and index of difficulty (*ID*). We compared the performance of the button to dwell using ray-based selection. The experiment was approved by our university ethics board.

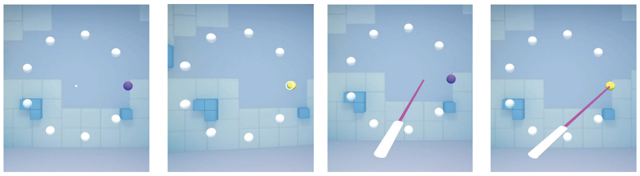
4.1 Participants

We recruited 28 participants (16F, 12M), aged 18-69 (M = 26.4, SD = 12.6 years) from our university and local community. All had (corrected-to-) normal vision, and all but one had tried VR at least once, with 45% stating they were novice to advanced users. Fifteen had no prior MVR experience and three had no experience with spatial input devices. Three were left-handed. Participants were compensated $15 CAD.

4.2 Apparatus

We used the following apparatus to conduct the experiments.

4.2.1 Hardware. We used the cardboard controller and the VRG Pro MVR HMD with a ~83° horizontal FOV and IPD adjuster. We



(a) (b) (c) (d)

Figure 2: Selection techniques before and during hover states. (a) Head-gaze cursor. (b) Head-gaze cursor hovering on target. (c) Controller and selection ray. (d) Ray hovering on target.

used a Samsung Galaxy S23 Ultra (Android OS 14). It 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 internal sensors provided 3DOF head rotations for viewpoint control.

4.2.2 Software. We developed a VR system in Unity 2021.3.12f1 using the Google Cardboard XR Plugin v.1.2. and Vuforia Engine v.10.15.3. The virtual environment was sourced from the Google Cardboard XR Plugin and sized to 6.72 × 6.72 × 4 meters. The camera (user's viewpoint) was positioned 1.7 m above the floor and only its orientation was changed by the smartphone's tracked rotations.

Head-gaze pointing used a ray originating at the camera position, and controlled by the camera orientation. A white dot cursor was displayed that expanded into a ring when hovering on an interactable object (i.e., the active target). See Figure 2b. The cursor was rendered in the target plane so cursor depth and stereo disparity were roughly the same as the targets themselves to prevent double vision and eye strain due to sudden convergence changes.

The controller ray originated at the centre of the three tracked markers. The ray direction was determined by the top markers' orientation, or the side markers' orientations if the top marker was not tracked in a given frame. The ray was displayed as a straight line emitted from the centre of a virtual controller, displayed as a white cylinder. Figure 2 shows both pointing techniques.

4.2.3 Task. The ISO selection task included nine evenly spaced spherical targets, positioned 75 cm in front of participants. All combinations of three target widths (25, 40, 55 cm) and amplitudes (35, 45, 60 cm) were used in the experiment, except the 40-45 combination to keep the experiment duration under an hour. While head and hand movements may cause the actual amplitudes and widths to vary in a 3D environment, these differences are negligible. Participants selected targets in sequence, with the current target coloured dark purple. Initially, the software displayed a start button; pressing it made the targets appear, and the system began recording data after the first selection (trial). The first target was selected twice in a round—once at the start, and once more at the end—yielding 10 trials per round (9 recorded). Sound data was only recorded between the first and final trials of each target ring.

For dwell selection activation, participants held the ray on a target for 300 ms to select it, following prior work [14]. During this dwell time, the target lightened in colour. Prematurely moving the ray off the target reset the target colour and timer. For button selection activation, the active target changed colours when the ray hovered on it. When the button press was registered, the active target changed to the inactive target colour, and the next target became active. Auditory feedback played upon each selection.

Missing targets was only possible with the button; all dwell trials ended with successful selection. To impose an accuracy requirement, selection was only possible in a region three times the target's radius. A hit required clicking the button while the ray touched the target; clicks missing the target but in the expanded region were recorded as misses. Clicks outside the region were ignored entirely.

4.3 Procedure

Participants first provided informed consent and demographic information. We then explained the experiment procedure, conditions, and how to use the controller and button. Participants sat in a swivel chair in a brightly lit and quiet lab environment while wearing the HMD. They first completed practice trials under all experiment conditions, before moving onto recorded trials. They were instructed to select targets as quickly and accurately as possible and could take breaks between any given round of trials. After all trials with a given selection activation method, participants completed the ISO Device Assessment Questionnaire [29] to solicit subjective preferences. Upon completing all conditions, a final questionnaire solicited overall preferences and feedback on the button's usability. The experiment lasted 60 minutes and participants were compensated upon completion.

4.4 Design

We employed a 2 × 2 × 8 within-subjects study design with the following independent variables:

- Selection Technique: Head-gaze, Controller
- Selection Activation: Dwell, Button
- ID: 2.37, 2.71, 2.94, 3.03, 3.35, 3.64, 3.73, 4.06

Our conditions included the four combinations of selection technique and activation method: Gaze-Dwell (*GD*), Gaze-Button (*GB*), Controller-Dwell (*CD*), and Controller-Button (*CB*). Eight *ID*s were repeated three times per condition, so participants completed 24 rounds of targets per condition with 9 trials (selections) each, for 864 recorded trials per participant. *ID* order was randomized, and condition order was counterbalanced via a balanced Latin square.

Dependent variables included selection time (ms), target reentries (count), selection distance (cm from target centre), errors (%), and throughput (bps). For the controller conditions (CD, CB), we recorded controller tracking data including tracking-loss duration (ms) and tracking-loss count. With the button conditions (GB, CB), we recorded audio data to determine audio recognition accuracy. If participants pressed the button, but at least one of the sound filters did not pass, a registration error was recorded. Any sound that did not pass all filters counted as registration errors. The quiet environment ensured errors were accurately recorded.

5 Results

We recorded 24,192 selections in total. Of those, 11 timed out at 30 s and 11 were discarded due to technical issues in data collection. Outliers were defined as ±3 *SD*s from the mean selection time of their respective condition. In total, 451 outliers (1.86% of 24,192) were removed, 209 (0.86%) from CB, 10 (.04%) from GD, 121 (0.5%) from GB,

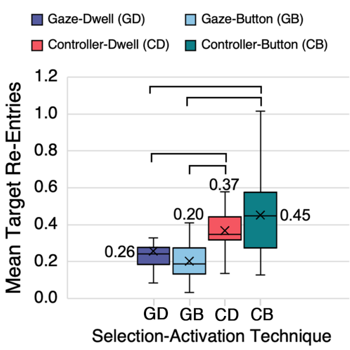


Figure 3: Mean target re-entries per trial by condition. Brackets show pairwise significance ($p < .05$).

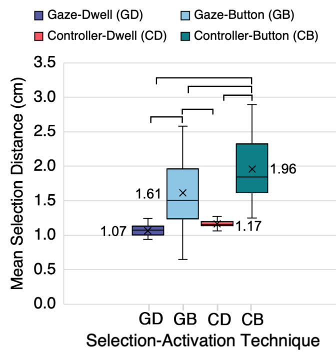


Figure 4: Mean selection distance by condition. Brackets show pairwise significance ($p < .05$).

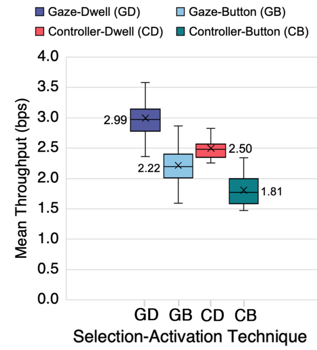
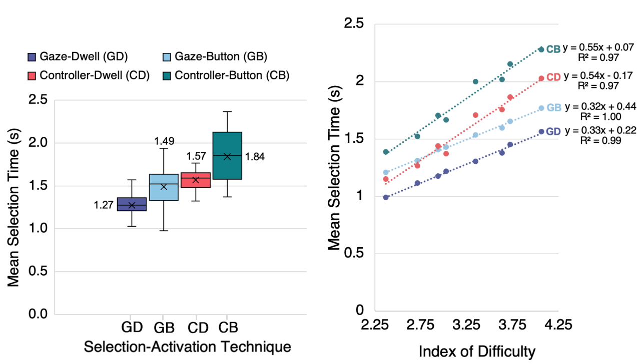


Figure 5: Mean Throughput scores by condition.

and 111 (0.46%) from CD. In total, 1.96% of the data was removed before analysis. A Lilliefors normality test on each dependent variable revealed all were normally distributed except tracking-loss duration. We analyzed the normally distributed metrics with two- or one-way ANOVA and used the Wilcoxon signed-rank test for the rest. Post-hoc analyses used Tukey HSD. There was no significant effect for condition order on selection time ($F_{3,24} = 1.75$, $p > .05$, $\eta_p^2 = .18$), suggesting counterbalancing was effective.

5.1 Selection Time

Selection time was the time in ms from a target activating to selection occurring. Two-way repeated measures ANOVA revealed a significant main effect for both selection technique ($F_{1,27} = 175.88$, $p < .001$, $\eta_p^2 = .87$, *pow.* = 1.0) and selection activation method ($F_{1,27} = 33.87$, $p < .001$, $\eta_p^2 = .56$, *pow.* = 1.0). Selection time was significantly lower with head-gaze pointing and dwell-based selection activation, as shown in Figure 6. No interaction effect was found ($F_{1,27} = 2.363$, $p = .14$, $\eta_p^2 = .08$, *pow.* = .2). Modelling the relationship between selection time and *ID*, showed all conditions were highly linear ($R^2 > 0.97$); Fitts' law is a strong predictor of selection time with both controller and button. See Figure 6b.



(a) **(b)**

Figure 6: Mean selection time by (a) condition and (b) *ID*.

5.2 Accuracy

For accuracy, we report target re-entries (*TRE*s) and selection distance; for button conditions (GB, CB) we also include error rates (i.e., selections outside the target but within 3 times the target radius). The mean error rate of GB and CB were 18% (*SD* = 12.76%) and 26% (*SD* = 12.76%) respectively. One-way ANOVA revealed a significant main effect ($F_{1,27} = 28.05$, $p < .001$, $\eta_p^2 = .51$, *pow.* = 1.0), indicating that head-gaze had lower error rates.

TREs indicate how many times the cursor/ray hit a target after the first hit; high TREs suggest control problems. Two-way repeated measures ANOVA revealed a significant main effect for selection technique on TRE ($F_{1,27} = 40.55$, $p < .001$, $\eta_p^2 = .60$, *pow.* = 1.0) and significant interaction between selection technique and activation method ($F_{1,27} = 12.78$, $p < .05$, $\eta_p^2 = .32$, *pow.* = .81). See Figure 3.

Finally, selection distance was the distance in cm from the target centre to the closest point on the selection ray. Two-way repeated measures ANOVA revealed the main effects for both selection technique ($F_{1,27} = 28.82$, $p < .001$, $\eta_p^2 = .55$, *pow.* = 1.0) and selection activation method ($F_{1,27} = 62.0$, $p < .001$, $\eta_p^2 = .70$, *pow.* = 1.0) were significant, as was the interaction effect ($F_{1,27} = 12.49$, $p < .005$, $\eta_p^2 = .32$, *pow.* = .08). See Figure 4. The large difference between button and dwell is likely due to the presence of errors with the button.

5.3 Throughput

We calculated *TP* using the equation $TP = ID_e/MT$, where $ID_e = log_2(A_e/W_e + 1)$. We use effective values to post-experimentally adjust W to an error rate of 4% and A to the actual distance the cursor moved for each selection to better represent the task performed. This decreases throughput's susceptibility to speed/accuracy tradeoffs [42, 64] and facilitates comparison of *TP* across studies with varying error rates.

Two-way repeated measures ANOVA showed significant main effects for both selection technique ($F_{1,27} = 102.83$, $p < .001$, $\eta_p^2 = .79$, *pow.* = 1.0) and selection activation method ($F_{1,27} = 326.48$, $p < .001$, $\eta_p^2 = .92$, *pow.* = 1.0). See Figure 5. No interaction effect was found ($F_{1,27} = 2.396$, $p = .13$, $\eta_p^2 = .08$, *pow.* = .2).

5.4 Tracking Performance

We also recorded how often tracking was lost (tracking-loss count) and for how long (tracking-loss duration). Both dependent variables were averaged by participant. A Wilcoxon signed-rank test indicated tracking-loss duration was significantly higher in CB than CD (z = -3.23, p = .001, r = -.43, *pow.* = .34). On average, the controller lost tracking for 226.41 ms with the button, compared to 117.52 ms with dwell. One-way ANOVA revealed a significant main effect of selection technique on tracking-loss count ($F_{1,27}$ = 7.59, $p < .05$, η_p^2 = .22, *pow.* = .77). On average, per trial, controller tracking was lost 0.11 times with button vs. 0.07 times with dwell.

5.5 Button Reliability

The software recorded audio clips passing all filters as a button click. False negatives were intentional clicks that were not registered while the selection ray was within 3 target radii (i.e., the distance used to define errors). Overall, the system had a 92.5% accuracy rate; 7.85% of clicks were false negatives. Error rate was the average of targets missed in each round of nine targets (7.49%). Overall, reliability was similar to acoustic sensing approaches, with accuracy rates of 90% [23] and 96% [77].

5.6 User Feedback

After completing a condition, participants filled out the ISO Device Assessment Questionnaire [29], excluding questions unrelated to selection activation. All questions were scored on a seven-point Likert scale with seven being most positive. As seen in Figure 7, dwell scored significantly higher except for fatigue and operation speed. Participants also ranked each condition. Results of this ranking are seen in Figure 8. We also asked, of dwell and button, which selection activation method they preferred; 20 (71.4%) participants preferred dwell, while 8 (28.6%) preferred the button.

We also collected subjective preferences at the end of the experiment. Participants who preferred dwell found its automation made it "*easier to advance in the trials quickly*" and found it less physically and mentally demanding as they did not need to "*coordinate [activating the selection while pointing]*". However, advantages of active vs. passive input [14, 16, 22, 46], may influence this opinion in real-world usage. Notably, three participants stated they would prefer the button if they did not have to perform many consecutive selections, as it caused their hand to "*hurt*" or start "*cramping*". One participant stated "*if the selections were more limited, I think I would have preferred the button*", while a second participant stated "*if [selections were] spread out more then [the] button would be ideal ... when you need a bit more precision*". Finally, four participants thought the button was less reliable since some clicks did not register, noting it was "*a bit too finicky*" and "*difficult to press down correctly*". This is consistent with our results on false negatives, and highlights the need for a more robust system.

Conversely, and consistent with past work [14], three participants who preferred the button noted greater control and accuracy, and that it prevented accidental selection. Three participants mentioned the button was more comfortable than waiting for the dwell timeout, with one stating "*keeping the head still pointing at the same point ... is challenging. The clicker let me move way faster*". Most participants who preferred the button felt it was faster, consistent with past work [46].

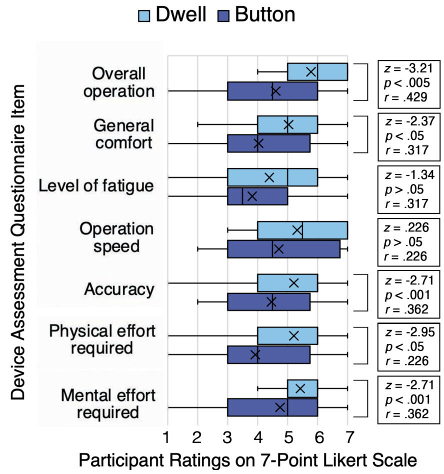


Figure 7: Device Assessment Questionnaire by selection activation. Central vertical lines indicate the median, X markers denote the mean. Brackets show pairwise significance.

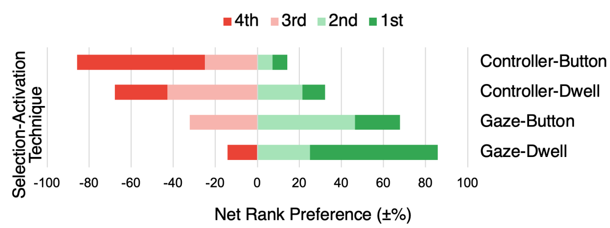


Figure 8: Participant rankings of conditions; total % of participants who chose each ranking.

6 Discussion

Overall, head-gaze was faster than the controller, and dwell outperformed the button, consistent with prior work [22]. As seen in Figure 6, GD was fastest, CB was slowest, and GB and CD were comparable. The controller's lower performance likely relates to tracking reliability; tracking loss was 117-226 ms on average, compromising smooth operation. Similarly, tracking was lost more often with the button. This may contribute to slower selection times and TP scores for CB. However, CB performance is in line with past controller-based selection with button input [65].

TP scores reveal a similar pattern: GD was highest (2.99 bps) and CB was lowest (1.81 bps). Notably, CD outperformed GB. This suggests selection activation had a higher impact on TP; for selection time, the selection technique had more influence. This may have occurred due to errors influencing TP and selection time for button conditions, since errors were not possible with dwell. Errors increase W_e, decreasing TP. The possibility of missing in the enlarged hit area also offered faster selection, as participants did not have to move the selection ray as far from the previous to current target for a selection attempt to register.

A key advantage of throughput is that its consistency facilitates comparison between studies better than movement time [64]. Computer mouse performance is a common baseline, at around 3.7-4.9 bps [64]. Generally, 3D pointing devices offer lower TP due to larger slower movements in 3D space. Pointing studies with comparable conditions report TP from 1.5-2.5 bps [22, 31, 56, 59, 65, 66]. Our controller with the button was in this range at 1.81 bps. This exceeds the TP of 1.06 bps reported in prior work using low-fi MVR controllers and marker-based selection activation [16]. This suggests that in the right conditions, a sound-based button can adequately support selection activation with low-fidelity 3D input devices.

In terms of accuracy, TREs were higher with the controller (see Figure 3). This may be due to the aforementioned Heisenberg Effect [33] or participant hand tremor and/or the use of optical tracking; the tracking notably introduced some input noise causing the ray to leave the target area. Notably, TREs were lower with GB than GD, despite GD outperforming GB in all other metrics. This may be due to participants having to hold their gaze longer with dwell, leaving more time for the cursor to leave the target area. Participant feedback supports this, indicating that GB offered both speed and control over selection, while GD caused discomfort, dizziness, and nausea due to fast head motions. As previously noted, the possibility of errors yielded a greater selection distance with the button than dwell. Missing a target was only possible with button conditions, increasing selection distance. When comparing button conditions, CB was significantly less accurate, consistent with TP scores.

To assess button reliability, we examined false negatives by selection technique; it was 4% (*SD* = 9%) with head-gaze vs. 8% (*SD* = 12%) with the controller. Of all registered clicks, 7.85% were false negatives. Most false negatives (65%) occurred with the controller. This may be due to how participants operated the button with each selection technique. With head-gaze, participants could hold the button stationary and keep their thumb at an optimal position. However, the controller required frequent re-adjustments of the thumb position, increasing false negatives and selection time.

Overall, the button offered comparable speed and TP to past studies with head-gaze [22, 56] and 3D pointing devices [65] with button activation. While dwell outperformed the button and was preferred, it is impractical for situations requiring greater control [14] and yields accidental selections in dense environments. Enhancing the sound recognition system and button design would improve overall performance, perhaps surpassing dwell. We discuss design improvements in the following section.

Based on our findings, we recommend optimal scenarios for each technique to guide researchers and practitioners. GB is best when accuracy is paramount, as it supports reliable, controlled selections. CB is preferable when users need greater input control, as decoupling selection from viewpoint control enables precise selection through fine-grained controller rotations to benefit navigation or search tasks [76]. For rapid selections and when fatigue is a concern, consider dwell time instead. User mobility also matters: head-gaze demands larger neck movements, while the controller and button require higher finger dexterity.

6.1 Design Improvements

Our goal is to provide a low-cost solution to 6DOF interaction in MVR that can be widely disseminated. Our results identified areas to improve the system design by increasing usability and robustness.

6.1.1 Physical Design. A purpose of our work was to provide a proof-of-concept for adding a button to a 6DOF controller. As such, the materials and button integration were rudimentary. Participants sometimes had difficulty making the sound if their thumb slid away from the optimal spot to press the keyring down onto the lid. This issue could be mitigated by adding tactile feedback such as a notch, nub, or texture on the interaction surface, similar to the ridges on a standard keyboard "f" and "j" keys. A simplified design better embedding the button into the controller could reduce the pieces and materials required, negating this issue altogether. A simplified, integrated design, could also support public distribution. For example, a DIY cardboard controller kit, akin to Nintendo's Labo[3], could include both button materials and pre-creased cardboard with printed markers, to be folded into a controller by end-users. Past work proposed plastic controllers for MVR [30]; however, our DIY approach offers advantages as it does not require hardware more expensive than the HMD, and the materials are easily sourced.

6.1.2 System Design. Machine learning could enhance system robustness without compromising performance in mobile VR environments, as prior work confirms [9, 77]. False positives from environmental noise and inconsistencies in user-generated sounds (e.g., pitch variations due to clicking force [63]) could be mitigated by adapting filter values to the environment and user when starting a session. Users could record a sample of button clicks that the system would use to adjust the filter values to effectively filter out noise from the environment and identify the user-made sound.

Adaptive filters could further enable user-defined sounds produced via arbitrary objects or surfaces. Past work shows user-defined surface gestures improved performance and learnability [9, 45, 61], suggesting advantages to giving users control over input actuation. This approach offers complete control over what materials and actions are used to make the sound. This would be highly convenient and enhance accessibility by supporting input customized to user's abilities. Particularly, users with limited hand or finger dexterity have difficulty operating VR controllers [10].

6.2 DIY for Emerging XR Platforms

The proposed system exemplifies how affordances of everyday objects can be repurposed for XR interactions. Our system leveraged the shape and sound of an object, but designers may consider all factors of everyday objects, such as their texture, reflectivity, or designed functionalities, and how they can be used to support different feedback modalities or other fundamental tasks; past work has utilized similar DIY practices to support selection, object manipulation, and travel [14–16, 18, 32, 68, 72, 73]. Beyond low-fi VR, DIY practice can be integrated with high-end VR systems to support custom input and rapid prototyping of input devices [28]. A more comprehensive discussion on DIY practices to support XR interaction is available elsewhere [20].

[3]https://www.nintendo.com/sg/switch/adfu/index.html

While an advantage of employing DIY practices is providing cost-accessible interactions, it can also support sustainable alternatives. Current cheap input devices, such as Bluetooth controls contribute to e-waste. Meanwhile, repurposing everyday objects for XR interactions can not only minimize environmental effects by not needing batteries and bring low-powered, but can actively reduce e-waste if old electronics (e.g., wired earbuds [73]) are repurposed.

Historically, Low-fi VR has shaped commercial hardware, bearing in mind the FOV2GO [49] inspired the Oculus Rift DK1. As such, when creative low-fi solutions are explored, they can inspire unique design approaches and provide valuable lessons for emerging XR platforms that may otherwise have remained unexplored.

6.3 Study Limitations

The experiment was conducted in a quiet environment to mitigate false positives, so all recorded errors were false negatives. Therefore, system robustness has not been thoroughly explored; registration accuracy may be worse in noisier environments. We also note the 300 ms dwell time is more appropriate for expert users [22, 43, 47] while a dwell time of 450 ms, more appropriate for novices [43], which may influence subjective results.

Nearly half of all outliers were with CB, suggesting issues with controller tracking and sound recognition increase dramatically when used together. During the experiment, we noticed a difference in how participants interacted with the button depending on the pointing technique. When paired with head-gaze, participants seemed to have an easier time with the button, resting it in their lap and continuously clicking. Controller tracking problems disrupted user flow and we noticed participants more frequently re-adjusted their thumb placement on the button. While this suggests the button does not pair well with the controller, improving integration between the button and controller may improve this.

6.4 Future Work

We plan to evaluate the button's efficacy in other commands, such as double click, or using multiple buttons emitting uniquely detectable sounds. This could increase the complexity of active user input for MVR, and thus support more interaction techniques common to high-end VR. With these added commands, we plan to evaluate the controller and button in interaction tasks beyond selection. This includes object docking and travel tasks. Unlike buttons, dwell cannot support multiple input commands. Thus, despite dwell's better performance, there is a need for button input in MVR. We proposed design iterations in section 6.1 that may improve button performance to be comparable or superior to a dwell time of 450 ms. Future work can explore how these updates affect the system's accuracy and user acceptance, and other possible advantages.

7 Conclusion

Mobile VR provides low-cost and broadly available immersive experiences, offering a potential solution to high-cost VR devices [26]. However, its poor interaction [16, 17, 33, 41, 54] and limited selection activation methods [14, 16, 41, 54] motivated our work on a low-cost unpowered button for MVR that emits clicks detected via acoustic sensing. An experimental comparison of the button to dwell indicates the button performance is similar to ray-based selection techniques used in prior studies [22, 56, 59, 65, 66] at 2.22 bps when used with head-gaze and 1.81 bps with a controller. The button had an overall accuracy rate of 92.5% with 7.85% of clicks registering as false negatives. These results validate the button as a viable input mechanism for 3D pointing techniques in MVR. Since dwell performed better overall, we plan to increase the system's robustness to improve button reliability. We proposed potential design updates to improve the button's usability and potential for public distribution as a cost-accessible input mechanism for MVR.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] Gadgets Now Bureau . 2023. Facing issues with gaming controller on your phone, here are 10 things to check. https://gadgetsnow.indiatimes.com/device-care/facing-issues-with-gaming-controller-on-your-phone-here-are-10-things-to-check/articleshow/106134016.cms

[2] Karan Ahuja, Rahul Islam, Varun Parashar, Kuntal Dey, Chris Harrison, and Mayank Goel. 2018. EyeSpyVR: Interactive Eye Sensing Using Off-the-Shelf, Smartphone-Based VR Headsets. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2 (7 2018), 10 pages. Issue 2. doi:10.1145/3214260

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

[4] Huidong Bai, Alaeddin Nassani, Barrett Ens, and Mark Billinghurst. 2017. Asymmetric bimanual interaction for mobile virtual reality. In *Proceedings of the 27th International Conference on Artificial Reality and Telexistence and 22nd Eurographics Symposium on Virtual Environments* (Adelaide, Australia) *(ICAT-EGVE '17)*. Eurographics Association, Goslar, DEU, 83–86. doi:10.2312/egve.20171343

[5] Simon Bothén, Jose Font, and Patrik Nilsson. 2018. An analysis and comparative user study on interactions in mobile virtual reality games. In *Proceedings of the 13th International Conference on the Foundations of Digital Games*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3235765.3235772

[6] Doug A Bowman and Larry F Hodges. 1999. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing* 10 (1999), 37–53. Issue 1. doi:10.1006/jvlc.1998.0111

[7] Doug A Bowman, Chadwick Wingrave, Joshua Campbell, and Vinh Ly. 2001. Using pinch gloves for both natural and abstract interaction techniques in virtual environments. In *Proceedings of HCI International* (New Orleans, Louisiana, USA). Lawrence Erlbaum, Mahwah, NJ, USA, 629–633.

[8] Mingshi Chen, Panlong Yang, Jie Xiong, Maotian Zhang, Youngki Lee, Chaocan Xiang, and Chang Tian. 2019. Your Table Can Be an Input Panel: Acoustic-based Device-Free Interaction Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3 (3 2019), 1–21. Issue 1. doi:10.1145/3314390

[9] 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 (2021), 2597–2607. Issue 5. doi:10.1109/TVCG.2021.3067689

[10] David M. Cook, Derani Dissanayake, and Kulwinder Kaur. 2019. Virtual Reality and Older Hands: Dexterity and accessibility in hand-held VR Control. In *Proceedings of the 5th International ACM In-Cooperation HCI and UX Conference* (Jakarta, Surabaya, Bali, Indonesia) *(CHIuXiD'19)*. Association for Computing Machinery, New York, NY, USA, 147-–151. doi:10.1145/3328243.3328262

[11] 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 (2020), 1–18. doi:10.1186/s41239-020-00228-9

[12] Mateus C L de Castro, J. P. de A Xavier, 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)*, Vol. 1. IEEE, Piscataway, NJ, USA, 895–900. doi:10.1109/ICIT46573.2021.9453573

[13] Panagiotis Drakopoulos, George alex Koulieris, and Katerina Mania. 2021. Eye tracking interaction on unmodified mobile VR headsets using the selfie camera. *ACM Transactions on Applied Perception (TAP)* 18 (2021), 1–20. Issue 3. doi:10.1145/3456875

[14] Chunming Gao. 2025. Study of Primitive Input Methods in Mobile Virtual Reality Systems for Educational Applications. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, Piscataway, NJ, USA, 1091–1097. doi:10.1109/CCWC62904.2025.10903741

[15] Scott W. Greenwald, Luke Loreti, Markus Funk, Ronen Zilberman, and Pattie Maes. 2016. Eye gaze tracking with google cardboard using purkinje images. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. Association for Computing Machinery, New York, NY, USA, 19–22. doi:10.1145/2993369.2993407

[16] Kristen Grinyer and Robert Teather. 2023. Low-Fi VR Controller: Improved Mobile Virtual Reality Interaction via Camera-Based Tracking. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (Sydney, NSW, Australia) *(SUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 5, 12 pages. doi:10.1145/3607822.3614517

[17] Kristen Grinyer and Robert J. Teather. 2024. Improving Inclusion of Virtual Reality Through Enhancing Interactions in Low-Fidelity VR. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (Orlando, FL, USA). IEEE, Piscataway, NJ, USA, 305–310. doi:10.1109/VRW62533.2024.00061

[18] Kristen Grinyer and Robert J. Teather. 2025. Cardboard Controller: A Cost-Effective Method to Support Complex Interactions in Mobile VR. *IEEE Transactions on Visualization and Computer Graphics* 31 (2025), 1–12. Issue 10. doi:10.1109/TVCG.2025.3581158

[19] Kristen Grinyer and Robert J. Teather. 2025. ClickSense: A Low-Cost Tangible Active User Input Method Using Passive Acoustic Sensing for Mobile Virtual Reality. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 136, 7 pages. doi:10.1145/3706599.3720000

[20] Kristen Grinyer and Robert J. Teather. 2026. A Systematic Review of Interaction Techniques for Mobile Virtual Reality. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (Barcelona, Spain) *(CHI'26)*. Association for Computing Machinery, New York, NY, USA, 25 pages. doi:10.1145/3772318.3791161

[21] 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*. Association for Computing Machinery, New York, NY, USA, 49–60. doi:10.1145/2984511.2984576

[22] 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*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3206343.3206344

[23] Chris Harrison and Scott E Hudson. 2008. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 205–208. doi:10.1145/1449715.1449747

[24] Chris Harrison, Julia Schwarz, and Scott E Hudson. 2011. TapSense: enhancing finger interaction on touch surfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 627–636. doi:10.1145/2047196.2047279

[25] 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*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3170427.3188629

[26] Kuo-Ting Huang, Christopher Ball, and Jess Francis. 2023. The Perceived Impacts of COVID-19 on Users' Acceptance of Virtual Reality Hardware: A Digital Divide Perspective. *American Behavioral Scientist* 69 (3 2023), 1155–1176. Issue 9. doi:10.1177/00027642231156775

[27] Yi-Jheng Huang, Kang-Yi Liu, Suiang-Shyan Lee, and I-Cheng Yeh. 2020. Evaluation of a Hybrid of Hand Gesture and Controller Inputs in Virtual Reality. *International Journal of Human–Computer Interaction* 37 (8 2020), 169–180. Issue 2. doi:10.1080/10447318.2020.1809248

[28] Muhammad I. M. Isham, Habibah N. Hj Haron, Farthan b. Mohamed, and Chan V. Siang. 2021. VR Welding Kit: Accuracy Comparison Between Smartphone VR and Standalone VR Using RMSE. In *2021 IEEE International Conference on Computing (ICOCO)*. IEEE, Piscataway, NJ, USA, 341–346. doi:10.1109/ICOCO53166.2021.9673577

[29] ISO/TC 159 Ergonomics. 2012. *Ergonomics of human-system interaction — Part 411: Evaluation methods for the design of physical input devices* (1 ed.). Standard. International Organization for Standardization. https://www.iso.org/standard/54106.html

[30] Hyouk Jang, Juheon Choi, and Gunhee Kim. 2019. POL360: A Universal Mobile VR Motion Controller using Polarized Light. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3359996.3364262

[31] 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 (3 2021), 484–495. Issue 5. doi:10.1080/10447318.2020.1826190 doi: 10.1080/10447318.2020.1826190.

[32] Kunihiro Kato and Homei Miyashita. 2015. Creating a Mobile Head-mounted Display with Proprietary Controllers for Interactive Virtual Reality Content. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. Association for Computing Machinery, New York, NY, USA, 35–36. doi:10.1145/2815585.2817776

[33] 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*. Association for Computing Machinery, New York, NY, USA, 365–366. doi:10.1145/2993369.2996311

[34] 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*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3485279.3485292

[35] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. 2017. *3D User Interfaces* (2 ed.). Addison-Wesley Professional, Boston, MA.

[36] Jihyun Lee, Byungmoon Kim, Bongwon Suh, and Eunyee Koh. 2016. Exploring the front touch interface for virtual reality headsets. *Conference on Human Factors in Computing Systems - Proceedings* 07-12-May-2016 (5 2016), 2585–2591. doi:10.1145/2851581.2892344

[37] Jialei Li, Isaac Cho, and Zachary Wartell. 2018. Evaluation of cursor offset on 3D selection in VR. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*. Association for Computing Machinery, New York, NY, USA, 120–129. doi:10.1145/3267782.3267797

[38] Pedro Lopes, Ricardo Jota, and Joaquim A Jorge. 2011. Augmenting touch interaction through acoustic sensing. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. Association for Computing Machinery, New York, NY, USA, 53–56. doi:10.1145/2076354.2076364

[39] Feiyu Lu and A Bowman Doug. 2021. Evaluating the Potential of Glanceable AR Interfaces for Authentic Everyday Uses. In *IEEE VR and 3D User Interfaces (VR)* (Lisboa, Portugal). IEEE, Piscataway, NJ, USA, 768–777. doi:10.1109/VR50410.2021.00104

[40] Gan Luo, Panlong Yang, Mingshi Chen, and Ping Li. 2020. HCI on the Table: Robust Gesture Recognition Using Acoustic Sensing in Your Hand. *IEEE Access* 8 (2020), 31481–31498. doi:10.1109/ACCESS.2020.2973305

[41] Siqi Luo, Robert J Teather, and Victoria McArthur. 2020. Camera-based selection with cardboard head-mounted displays. In *HCI International 2020–Late Breaking Papers: Virtual and Augmented Reality* (Copenhagen, Denmark) *(HCII 2020)*. Springer, Berlin, Heidelberg, 383–402. doi:10.1007/978-3-030-59990-4_29

[42] 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*. Association for Computing Machinery, New York, NY, USA, 1633–1636. doi:10.1145/1357054.1357308

[43] 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* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. doi:10.1145/1518701.1518758

[44] 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*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3290605.3300815

[45] Meredith Ringel Morris, Jacob O Wobbrock, and Andrew D Wilson. 2010. Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010*. Association for Computing Machinery, New York, NY, USA, 261–268.

[46] 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*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3448018.3457998

[47] Aanand Nayyar, Utkarsh Dwivedi, Karan Ahuja, Nitendra Rajput, Seema Nagar, and Kuntal Dey. 2017. OptiDwell: Intelligent Adjustment of Dwell Click Time. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 193–204. doi:10.1145/3025171.3025202

[48] Kent L. Norman and Jurek Kirakowski (Eds.). 2018. *The Wiley Handbook of Human Computer Interaction* (1 ed.). Vol. 1. Wiley, Hoboken, NJ, USA. 349–370 pages. doi:10.1002/9781118976005

[49] J. Logan Olson, David M. Krum, Evan A. Suma, and Mark Bolas. 2011. A design for a smartphone-based head mounted display. In *IEEE Virtual Reality Conference*. IEEE, Piscataway, NJ, USA, 233–234. doi:10.1109/VR.2011.5759484

[50] Kyeong-Beom Park and Jae Yeol Lee. 2019. New design and comparative analysis of smartwatch metaphor-based hand gestures for 3D navigation in mobile virtual reality. *Multimedia Tools and Applications* 78 (3 2019), 6211–6231. Issue 5. doi:10.1007/s11042-018-6403-9

[51] Pew Research Center. 2025. *Mobile Fact Sheet*. Technical Report. Pew Research Center, Washington, D.C. https://www.pewresearch.org/internet/fact-sheet/mobile/

[52] 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)*. IEEE, Piscataway, NJ, USA, 172–175. doi:10.1109/3DUI.2017.7893335

[53] Ivan Poupyrev, Tadao Ichikawa, Suzanne Weghorst, and Mark Billinghurst. 1998. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum* (Oxford, UK and Boston, USA), Vol. 17. Blackwell Publishers Ltd, Oxford, UK, 41–52. Issue 3. doi:10.1111/1467-8659.00252

[54] 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)*. IEEE, Piscataway, NJ, USA, 5–8. doi:10.1109/WEVR.2016.7859536

[55] Kevin Purdy. 2016. Understanding Bluetooth Pairing Problems. https://www.nytimes.com/wirecutter/blog/understanding-bluetooth-pairing-problems/

[56] 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* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. doi:10.1145/3131277.3132182

[57] Massimo Ragnedda and Hanna Kreitem. 2018. The three levels of digital divide in East EU countries. *World of Media. Journal of Russian Media and Journalism Studies* 1 (12 2018), 5–26. Issue 4. doi:10.30547/WORLDOFMEDIA.4.2018.1

[58] Massimo Ragnedda and Maria Laura Ruiu. 2017. *Social capital and the three levels of digital divide* (1 ed.). Routledge, New York, NY, USA, Chapter 2, 1–14. doi:10.4324/9781315455334

[59] Adrian Ramcharitar and Robert J Teather. 2018. EZCursorVR: 2D Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 44th Graphics Interface Conference*. Association for Computing Machinery, New York, NY, USA, 123–130. doi:10.20380/GI2018.17

[60] Simon Robinson, Nitendra Rajput, Matt Jones, Anupam Jain, Shrey Sahay, and Amit Nanavati. 2011. TapBack: towards richer mobile interfaces in impoverished contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2733–2736. doi:10.1145/1978942.1979345

[61] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 197–206. doi:10.1145/1978942.1978971

[62] Valkyrie Savage, Andrew Head, Björn Hartmann, Dan B Goldman, Gautham Mysore, and Wilmot Li. 2015. Lamello: Passive Acoustic Sensing for Tangible Input Components. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1277–1280. doi:10.1145/2702123.2702207

[63] Lei Shi, Maryam Ashoori, Yunfeng Zhang, and Shiri Azenkot. 2018. Knock knock, what's there: converting passive objects into customizable smart controllers. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3229434.3229453

[64] 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 (2004), 751–789. Issue 6. doi:10.1016/j.ijhcs.2004.09.001

[65] 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)* (Singapore). IEEE, Piscataway, NJ, USA, 87–94. doi:10.1109/3DUI.2011.5759222

[66] 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*. Association for Computing Machinery, New York, NY, USA, 159–168. doi:10.1145/2470654.2470677

[67] 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* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 417–427. doi:10.1145/3332165.3347946

[68] Anran Wang and Shyamnath Gollakota. 2019. MilliSonic: Pushing the Limits of Acoustic Motion Tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3290605.3300248

[69] AJ Wykes. 2021. Why is Bluetooth still so unreliable? https://www.soundguys.com/why-is-bluetooth-unreliable-55016/

[70] Robert Xiao, Greg Lew, James Marsanico, Divya Hariharan, Scott Hudson, and Chris Harrison. 2014. Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*. Association for Computing Machinery, New York, NY, USA, 67–76. doi:10.1145/2628363.2628383

[71] 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* (Brisbane, QLD, Australia) *(DIS '16)*. Association for Computing Machinery, New York, NY, USA, 229–233. doi:10.1145/2901790.2901813

[72] Majed Al Zayer and Eelke Folmer. 2018. StereoTrack: 180-degree Low-cost Acoustic Positional Tracking for Mobile VR Platforms. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. Association for Computing Machinery, New York, NY, USA, 143–154. doi:10.1145/3270316.3273044

[73] Majed Al Zayer, Sam Tregillus, and Eelke Folmer. 2016. PAWdio: Hand Input for Mobile VR using Acoustic Sensing. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, New York, NY, USA, 154–158. doi:10.1145/2967934.2968079

[74] Cheng Zhang, Anandghan Waghmare, Pranav Kundra, Yiming Pu, Scott Gilliland, Thomas Ploetz, Thad E Starner, Omer T Inan, and Gregory D Abowd. 2017. FingerSound: Recognizing unistroke thumb gestures using a ring. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3 (9 2017), 1–19. Issue 3. doi:10.1145/3130985

[75] Daniel Zielasko, Sebastian Freitag, Dominik Rausch, Yuen C Law, Benjamin Weyers, and Torsten W Kuhlen. 2015. BlowClick: A Non-Verbal Vocal Input Metaphor for Clicking. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. Association for Computing Machinery, New York, NY, USA, 20–23. doi:10.1145/2788940.2788953

[76] Daniel Zielasko, Yuen C. Law, and Benjamin Weyers. 2020. Take a Look Around – The Impact of Decoupling Gaze and Travel-direction in Seated and Ground-based Virtual Reality Utilizing Torso-directed Steering. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA, USA). IEEE, Piscataway, NJ, USA, 398–406. doi:10.1109/VR46266.2020.00060

[77] Daniel Zielasko and Javier Alejandro Jaquez Lora. 2024. Ubiquitous BlowClick: Non-verbal Vocal Input for Confirmation with Hand-held Mobile Devices in the Field. In *Proceedings of the 2024 ACM Symposium on Spatial User Interaction*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3677386.3682101

Received 3 April 2026