RESEARCH-ARTICLE

Beyond Controllers: Exploring On-Body Micro-Gestures for Mixed Reality Adventure Gameplay

MATILDA HANELING, KTH Royal Institute of Technology, Stockholm, Stockholms, Sweden

JESPER LUNDQVIST

ANDRII MATVIIENKO, KTH Royal Institute of Technology, Stockholm, Stockholms, Sweden

Published: 12 November 2025

VRST '25: 31st ACM Symposium on Virtual Reality Software and Technology
November 12 - 14, 2025
QC, Montreal, Canada

Conference Sponsors:
SIGCHI
SIGGRAPH

VRST '25: Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology (November 2025)
https://doi.org/10.1145/3756884.3766008
ISBN: 9798400721182

Beyond Controllers: Exploring On-Body Micro-Gestures for Mixed Reality Adventure Gameplay

Matilda Haneling
KTH Royal Institute of Technology
Stockholm, Sweden
haneling@kth.se

Jesper Lundqvist
Resolultion Games
Stockholm, Sweden
jesper.lundqvist@resolutiongames.com

Andrii Matviienko
KTH Royal Institute of Technology
Stockholm, Sweden
andriim@kth.se

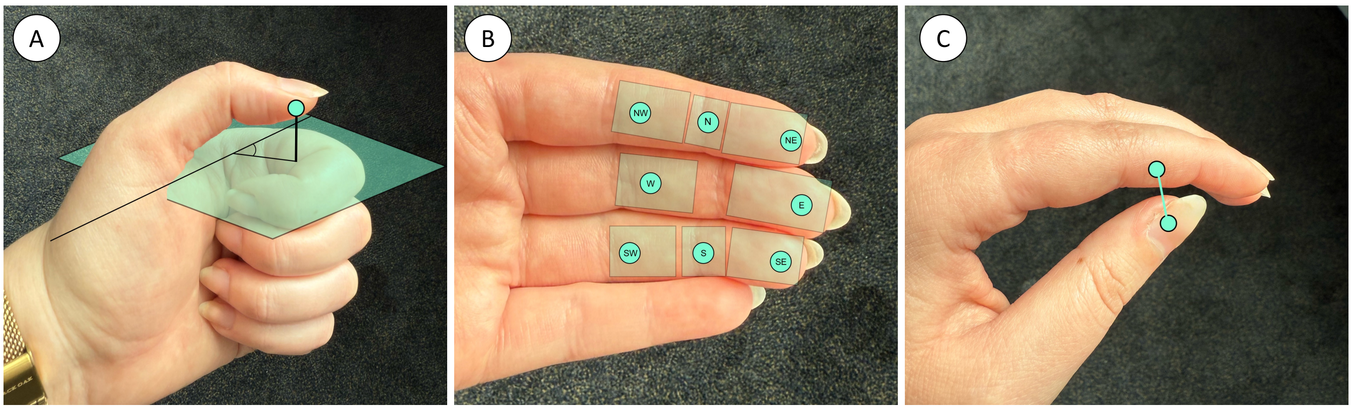

Figure 1: (A) Knuckle: The thumb's position relative to a projected plane on the hand determines input direction via the angle from the center. (B) Palm: Input direction is based on the thumb's distance to predefined palm joint areas; if it falls within a threshold of one area (C), that area defines the direction.

Abstract

Mixed-reality (MR) platforms are moving toward controller-free input, relying on hand gestures and eye tracking. In this paper, we explore on-body micro-gestures as alternatives to MR gameplay via two implemented techniques – *Palm* and *Knuckle* – modeled on joystick steering and button tapping. We evaluated them with 24 participants, who played a third-person MR adventure game using the proposed techniques and a gaming controller. We found that the controller leads to the fastest and most precise performance. While *Palm* produced the most attacks and enemy kills and imposed the highest workload, *Knuckle* was rated the most enjoyable and socially acceptable in public. We demonstrate that although physical controllers are superior in speed and accuracy, on-body micro-gestures emerge as an appealing alternative/complement for casual or mobile MR play.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; **Gestural input**; **Empirical studies in interaction design**.

Keywords

mixed reality, on-body, micro-gestures, hand tracking, games

VRST '25, Montreal, QC, Canada

ACM ISBN 979-8-4007-2118-2/25/11
https://doi.org/10.1145/3756884.3766008

ACM Reference Format:
Matilda Haneling, Jesper Lundqvist, and Andrii Matviienko. 2025. Beyond Controllers: Exploring On-Body Micro-Gestures for Mixed Reality Adventure Gameplay. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3756884.3766008

1 Introduction

Mixed Reality (MR) is pivotal on the extended-reality spectrum, blending co-present virtual and physical elements in real time [41]. The 2024–2025 cycle marks a decisive hardware shift: Apple's *Vision Pro* launched with an interface "controlled entirely by a user's eyes, hands, and voice," deliberately omitting physical controllers [2]. Samsung's upcoming Android XR headset follows the same paradigm, relying primarily on hand- and eye-tracking [23]. While this trend promises more natural, "hands-free" interaction, it introduces two practical drawbacks. Mid-air gestures without the support of a physical object lead to user fatigue over extended sessions [22, 24], and they also lack the tactile and haptic feedback that physical controllers rely heavily on. One area where this has a huge impact is gaming, where tactile controllers have been a key component and highly integrated in games since the early 1970s [8] and have more or less been standardized in terms of functionality. While the gaming industry usually does not tend to be scared to innovate, it has even been noted that specific controls appear to be one of the most challenging parts to make significant changes [8]. On-body interaction as an alternative to controller-based interaction in gaming contexts presents an interesting research opportunity for gaming, which we explore in this paper.

Prior research on gesture-based input in MR leaves three persistent gaps. First, many studies employ bigger, whole-arm or mid-air gestures whose cinematic flair comes at the cost of physical fatigue and limited precision [3, 9, 37]. Second, micro-gestures, defined as physical movements of fingers, which are recognized by the system, and where the system reacts upon [45], are often considered for a single task or device rather than general-purpose game control—examples range from chess-piece pinching [5] to rehabilitation drills [25]. Finally, many studies opt for more extensive setups, with specially designed gloves or other tactile proxies, which directly contradicts the previously mentioned goal of more natural and "hands-free" embodiment. Comparative evaluations consistently reveal that bare-hand input enhances presence, embodiment, or intrinsic motivation yet still lags behind controllers on speed, error rate, or workload in fast or accuracy-critical contexts [1, 18, 28, 30, 32]. Consequently, we lack an understanding of using localized, game-oriented micro-gestures that fit within the proprioceptive envelope of the hands and map cleanly onto familiar controller functions, without external hardware.

In this paper, we investigate how hand-based micro-gestures as an input influence user performance and experience compared to game controllers in mixed-reality adventure gameplay. Grounded in Norman's principle of consistency [35] and the Hand-Proximate UI framework [12], the gesture set mirrors two canonical controller primitives: **continuous steering** (thumb circling on the opposite palm or knuckle to emulate an analogue stick) and **discrete triggering** (thumb–finger taps that yield tactile "clicks" analogous to face-buttons). We designed and developed two types of on-body gestures – *Knuckle* and *Palm* – and compared them to a physical game controller in a study (N=24), in which participants played a third-person MR treasure hunt adventure game since it requires frequent use of micro-gestures while focusing on the primary task of gaming. *Knuckle* and *Palm* are intra-manual, i.e., performed within the hand's own space, making them less visually distracting and more compatible with MR scenarios where users must balance gameplay with awareness of the physical environment. Our findings showed clear trade-offs: the controller delivered the fastest completion times and was rated most intuitive and precise; Palm micro-gestures prompted the highest attack and kill counts but imposed the most significant physical workload; Knuckle micro-gestures matched the controller on movement efficiency and were judged the most enjoyable and socially acceptable in public.

2 Related work

2.1 Gestures in Immersive Spaces

Early camera-based tracking showed that bare-hand input could control virtual and MR environments. Rautaray et al. [37] demonstrated that a single camera could detect hand movements in real time for in-game actions. Chan et al. [9] expanded this with Germane, using a low-cost webcam to recognize gestures and manipulate virtual objects. To improve stability, Wilk et al. [44] paired a data glove with a head-mounted display, achieving reliable finger-level control in immersive gameplay. To restore tactile sensation, researchers introduced physical proxies [27]. Nilsson et al. [34] found that haptic cues improved interaction when a proxy matched the virtual object's properties, though practical only for tasks with feasible replicas. Feick et al. [13] developed VoxelHap, a toolkit for designing task-specific proxies, which users sometimes preferred over standard VR controllers. Saint-Aubert et al. [38] used a tangible puppet controller, resulting in higher ratings for fun, presence, and embodiment than a traditional D-pad. Controlled studies comparing hand gestures and controllers show mixed results. Hameed et al. [18] found controllers yielded faster times, fewer errors, and lower mental workload in motor tasks. Yet Juan et al. [25] observed higher motivation and acceptance for hands-free rehabilitation, despite faster task completion with controllers. Voigt-Antons et al. [43] reported higher valence and realism with hand tracking but lower dominance; most users still preferred controllers. Venkatakrishnan et al. [42] noted that rich hand models increase embodiment, but mismatches can harm performance. Lin et al. [28] showed that while virtual-hand illusions are strongest with direct hand control, handheld devices remain superior for precision tasks. Similarly, Moehring and Fröehlich [32] found finger input more realistic, but controllers faster and more reliable without added tactile feedback. Luong et al. [30] reported controllers were better for ray-cast pointing, while hands were preferred for mid-air touch tasks. These studies highlight a trade-off: bare-hand input enhances presence and emotional engagement, while controllers lead in speed, accuracy, and endurance. Hamari et al. [17] note that utilitarian game use depends on usefulness, while hedonic games are driven by enjoyment. This makes games a compelling context in which to explore the benefits of hand tracking.

2.2 Gestures in Games

Gesture-based interaction in MR gaming has been widely studied, often in comparison to traditional controllers. While controllers remain prevalent, researchers increasingly explore hand-tracking and gesture recognition as more natural, intuitive, and immersive alternatives. Prior work highlights these systems' strengths and limitations, underscoring their potential for controller-free interaction in MR gaming. A key advancement is AuroraSync by Sharma et al. [39], which uses real-time tracking of 21 hand landmarks to enable device control and interaction through natural gestures. By removing the need for controllers, AuroraSync reduces hardware dependency while preserving control and offering benefits like lower costs, simpler setup, and greater immersion. Similarly, Babu et al. [3] examine gesture-based interfaces across domains, including gaming, showing how they enhance accessibility and usability through more direct and instinctive interaction. Gesture-based interactions in gaming have been explored through various application-specific studies. Bikos et al. [5] developed an AR chess game using pinch gestures to manipulate virtual pieces. This shows that simple, natural gestures enhance immersion by bridging real-world actions and in-game interactions. Building on this, Geetha et al. [15] found that hand-tracking in MR environments improves interaction, control, and presence by enabling seamless and intuitive engagement with virtual content. Chan et al. [9] introduced Germane, a low-cost, camera-based system for gesture recognition, demonstrating that even basic hardware can offer a viable alternative to traditional controllers. In an escape-room game study, Adkins et al. [1] found that continuous hand tracking increased ownership, realism, enjoyment, and presence without reducing task efficiency. Visibility of virtual hands during grasps further enhanced these effects, suggesting that visual feedback can compensate for the lack of haptics. Taken together, these game-focused studies reaffirm the tension relationship between gestures and raise key challenges for the success of gesture-based systems in games. Its practical value hinges on reliable recognition, low fatigue, and clear visual feedback to offset missing tactile feedback. While hardware constraints limit recognition, this study tries to reduce user fatigue through micro-gestures.

2.3 On-Body Interaction

On-body interaction enables users to perform gestures on their own body—like tapping the palm, swiping the forearm, or pinching fingers—to engage with digital interfaces. Unlike traditional controllers, it offers a direct, embodied experience aligned with natural movement and tactile awareness [6, 10, 11]. Foundational work such as Skinput by Harrison et al. [20] demonstrated that the skin could serve as an input surface by detecting vibrations from finger taps, providing a portable and always-available method. They later introduced Armura [19], integrating input and graphical output onto the user's body. Previous studies have also shown cognitive and experiential advantages. For example, Bergstrom-Lehtovirta et al. [4] found that skin-based input enhances the sense of agency compared to button presses and touchpads. Similarly, Gustafson [16] showed that tactile and visual cues from the body support palm-based input as a viable alternative to touchscreens. A key study, EgoTouch by Mollyn and Harrison [33], uses AR/VR headset cameras to track on-body interactions—like touches on the palm or forearm—without extra sensors. It shows how on-body input can effectively replace controllers in MR environments. Given MR's emphasis on spatial awareness and intuitive interaction, on-body gestures offer a seamless and immersive alternative to external controllers. Building on the previous work, this study prioritizes on-body input over mid-air gestures, as it better compensates for the lack of tactile feedback and explores its suitability for mixed reality adventure gameplay as a potential alternative for controllers.

2.4 Micro-Gestures

Micro-gestures are subtle finger or hand movements—such as tapping, pinching, or twitching—that require minimal effort and are often performed unconsciously [45]. Unlike larger, traditional gestures, they offer intuitive, low-effort, and discreet interaction, especially valuable in constrained or attention-sensitive contexts like driving, AR, and wearables [29, 36, 46]. Advancements in sensing technologies have increased interest in micro-gestures as natural and seamless input methods. Peng et al. [36] developed a recognition system using an attention-based neural network, achieving high accuracy for AR/VR applications. Liu et al. [29] created the HoMG database to support 3D gesture recognition in wearable contexts. Gao et al. [14] showed that unintentional micro-gestures reveal emotional states, aiding emotional AI. Xiao et al. [46] explored car micro-gestures, enabling safe, intuitive in-vehicle interactions. These studies show that on-body micro-gestures can serve as effective MR input methods. However, key challenges remain: (1) lack of haptic feedback, (2) fatigue, and (3) recognition accuracy. This paper explores gesture-based interaction concepts using on-body micro-gestures for gameplay, detailed in the next section. These developments underscore the relevance of micro-gestures for this study, particularly as an alternative input method in mixed-reality environments. By examining how hand-based micro-gestures influence user performance and experience compared to traditional game controllers, this research explores whether these subtle interactions can offer more intuitive, efficient, or enjoyable gameplay in adventure-based mixed-reality scenarios.

3 Hand Gestures as Input for Gameplay

The main idea behind the gesture concept is Don Norman's principles of consistency and conceptual models. When an unfamiliar interface behaves like a familiar one, cognitive load falls and learnability rises [35]. As a result, the gesture set deliberately mirrors the two dominant control primitives of a standard controller:*Continuous steering* (the left joystick that moves the avatar) and *Discrete triggering* (the cluster of right-hand face-buttons that fire actions). This familiar split lets experienced players bring over their muscle memory from traditional controllers. It gives less experienced players a clear mental model, separating the input mode (continuous vs. discrete) and keeping actions separated between different hands (left for movement, right for actions). Additionally, the gestures are grounded in the Hand-Proximate User Interface (HPUI) framework introduced by Faleel et al. [12]. The framework proposes using the hands and the immediate space around them to register virtual interfaces. Hence, interaction occurs in a zone of rich proprioceptive and tactile feedback, further highlighting the suitability of using the hands and fingers for this input type. For locomotion, the left thumb is used. The thumb traces tight circles (< 20 mm in diameter) while resting lightly on the skin of the left hand. Because the hand itself never leaves a relaxed, elbows-in posture, or even keeps hands on the knees or the table, the motion is intended to feel close to moving an analogue stick while still staying ergonomic and non-strenuous.

Action inputs are kept on the right-hand side to mimic the buttons on the right-hand side of the controller. Using the right thumb, the tips of the remaining four fingers on the same hand act as easily accessible areas for fast actions. To replicate the feeling of pressing a physical button, the user would quickly tap the thumb and one of the fingertips together. Because the thumb and finger tips come into direct contact, users get a clear tactile "click" that mid-air gestures lack. Similarly to the left-hand actions, the user can keep their hand in a natural position, largely free to choose, as the distinct gesture should be easy to pick up for the gesture recognition system. The same gesture idea can also be reused for multiple actions, but using different finger tips, reducing the number of gestures the user has to learn and remember.

This pinch-based mapping preserves the controller's logic (right-hand for buttons, left-hand for movement) while replacing plastic hardware with the player's skin, keeping the vocabulary compact and easy to learn. Similar to a controller, these gestures can be used for different actions depending on the game's requirements, keeping the gestures general enough for multiple actions in different gaming contexts. This way, users would not have to learn new gestures for each new game. For this study, this concept has been realized in two technical implementations for the left hand, which handles

Figure 2: Image of how the game looks from the player's point of view (left) and simulated image of how the game was positioned during the test (right).

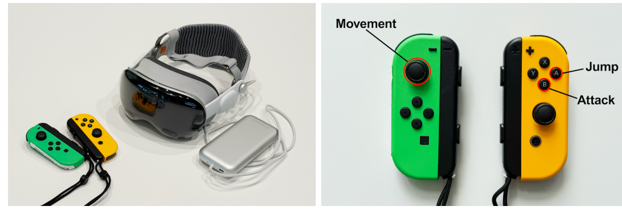

Figure 3: The hardware used in the user tests. Apple Vision Pro headset and Nintendo JoyCons (left) and the control scheme of the Controller condition (right).

movement: a palm-steering and a knuckle-steering, while the right-hand control remains unchanged.

4 Evaluation

The goal of our study was to compare the three input methods for Mixed Reality adventure gameplay with the following research question: *How do knuckle- and palm-based micro-gestures as an input influence user performance and experience compared to game controllers in a puzzle-oriented task focused on locomotion and trigger use in a reachable tabletop mixed-reality adventure gameplay?*

4.1 Participants

We recruited 24 participants (8 F, 16 M) aged between 22 to 56 years ($M = 29.5, SD = 8.7$). Most (20 participants) said they play games regularly (= at least monthly). However, most participants had limited experience with VR and AR headsets, and 13 participants had tried it a few times. As for the gestures, a majority (17 participants) marked "I have tried it a few times" using the Nintendo Wii console as an example. Three participants had experience with hand tracking in MR.

4.2 Study Design and the Game

This study was designed as a within-subjects study with one independent variable: *input method* – Controller, Knuckle, and palm. The Controller is a baseline, and Knuckle and Palm are based on the descriptions in the previous section. Each method reflects one experimental condition, and all conditions were counterbalanced using a Balanced Latin Square.

We developed a third-person adventure game using Unity (version 2022.3.21f), which consisted of 13 small scenes and areas, including a training area before the actual game started. These areas included a small house, a garden, the woods, and a castle with a dungeon. To finish the game, the player had to collect 20 bottles scattered throughout the different areas, a task given to the player by a non-playable character (NPC), from hereon referred to as "the wizard". After completing the task and returning the bottles to the wizard, the player was awarded a key to open the door to the winning treasure room. The game logic was used to make the players extensively use the controls, running back and forth between the different areas using eight directions (N, NE, E, SE, S, SW, W, NW). The bottles were placed at a height that forced the player to jump to reach them. In some scenes, enemies in the form of skeletons were placed. If the player got too close to these, they would start to attack. The player could then use their sword to attack and eventually kill these skeletons, which would respawn after a few seconds. However, the player would take no damage from the enemies, crashes, or falls, and could not die during the game. This was to make the game easier, especially for people less comfortable in a gaming setting, and to let players concentrate on the controls rather than staying alive during the game.

4.3 Measures

We measured the following dependent variables:

- **Task Completion Time (s)**: time participants took to finish the task, i.e., successfully collecting all 20 bottles.
- **Number of Jumps, Attacks, enemies killed, direction changes**: the number of times participants triggered the Jump & Attack actions, number of enemies killed, and direction changes.
- **Distance Moved (units)**: the horizontal distance the character moved in-game.
- **Perceived Workload**: after each condition, the participants rated their perceived workload using the NASA TLX [21].
- **User Experience Ratings**: participants rated the intuitiveness input methods, the ease of learning, the feeling of control, how well they felt they performed, fun and engagement, the tracking and recognition accuracy, how they felt about the long-term usability, how well they enjoyed the tactile feedback, and to what extent they felt the method could be used outside of gaming contexts.

4.4 Apparatus and Implementation

The game was deployed to an Apple Vision Pro headset on VisionOS 2.0. We used Nintendo JoyCons as a VR controller paired to the headset via Bluetooth, following conventional patterns of the left analog stick for movement and the buttons on the right D-pad for the two actions: Jump and Attack (Figure 3). To ensure a fair comparison, these actions were implemented similarly for Knuckle and Palm methods, preventing jump+attack concurrency.

Knuckle Method. At every frame, we retrieve six left-hand joint poses supplied by Unity XR Hands: three along the index finger (proximal, intermediate, distal) and the wrist transform. The three index-finger joints uniquely define a finger plane via the standard three-point plane equation. This dynamic plane follows finger motion and approximates the dorsal surface of the curled finger. Mapping all relevant points onto the finger plane reduces the problem from 3D pose estimation to 2D vector geometry, simplifying direction classification and eliminating any wrist pitch or roll effect. The thumb tip, thumb distal joint, index-finger tip, and index-finger distal joint are orthogonally projected onto the finger plane (i.e., each point's shortest-distance "shadow" is calculated, Figure 1). For robustness, two intermediate mid-points are formed: (1) Thumb mid-point – equidistant between projected thumb tip and distal, (2) Index mid-point – a weighted centroid (65% closer to the tip). These mid-points provide stable anchors representing overall thumb and finger positions better than any single joint, compensating for small skeletal-tracking noise. The system treats the gesture as "active" only when the thumb approaches the index finger within 22 mm (press threshold). Release occurs at $24mm$. The deliberate gap between the two thresholds constitutes a hysteresis band, preventing rapid on–off flicker due to micro-movements [1]. While active, an additional hand-shape must be satisfied to ensure the user is intentionally forming a D-pad gesture and not merely brushing fingers together: (1) distal-distal separations ($<$ 30 mm) and (2) tip-to-metacarpal distances – ($<$ 10 mm). Only when these geometric constraints hold is the D-pad considered "armed". The filter ensures that normal pointing or grasping postures do not trigger locomotion. During an active pinch the 2D displacement vector $\mathbf{d} = (x_{\text{thumb}} - x_{\text{index}},\ z_{\text{index}} - z_{\text{thumb}})$ is evaluated on the finger plane. Magnitude constraints are applied: (1) Activation radius: $>$ 55 mm to ignore negligible motion, (2) Outer discard radius: $>$ 60 mm to reject implausible excursions. If valid, the vector is normalized, its polar angle extracted via $\tan^{-1}(y/x)$, and the circle partitioned into eight 45° sectors, e.g., N, NE, E. The sector label becomes the discrete D-pad direction, and the normalized vector (±1 on each axis) is forwarded to downstream gameplay logic.

Palm Method. The Palm method lets the player "press" an eight-way D-pad against the palm side of the curled fingers. Unlike the Knuckle, it relies exclusively on local distances between the thumb tip and selected skeletal landmarks, making it insensitive to wrist roll or forearm pronation. The Unity XR Hands subsystem streams the full left-hand skeleton for every frame. The algorithm draws on the thumb tip plus three joints per index-, middle-, and ring-finger (intermediate, distal, tip). To stabilize these raw positions, each joint is averaged over the most recent fifteen frames; afterwards, one derived point per finger is calculated:

$$\mathbf{M}_f = \frac{1}{2}(\mathbf{I}_f + \mathbf{D}_f), \qquad f \in \{\text{index}, \text{middle}, \text{ring}\},$$

where $\mathbf{I}_f$ and $\mathbf{D}_f$ denote the intermediate and distal joint positions of finger f, respectively. Figure 1 schematically shows all ten target points that will later serve as points and areas for hit-testing. Similarly to the Knuckle method, before any directional logic is applied, the hand must adopt an armed posture: (1) the distal tips of index-to-middle and middle-to-ring fingers are closer than 30 mm, i.e the fingers cannot be spread out, (2) simultaneously, all three fingers cannot ne too curled toward the palm (tip–metacarpal distance > 10 mm). Once armed, the thumb tip is compared against six spatial rails and three joint positions. For each area, the algorithm projects the thumb onto a line segment $\overline{AB}$. It verifies two conditions: (1) The projection falls inside the segment ($\mathbf{0} \leq \lambda \leq \mathbf{1}$), (2) The Euclidean thumb–rail distance is below a direction-specific threshold, where the left-facing directions have a slightly higher threshold, as they, in preliminary testing, proved harder or more straining to reach. With a higher threshold, these directions become more forgiving and easier to reach. The D-pad reverts to neutral if none of the areas satisfy these two conditions. Otherwise, the stick vector is generated. A single-point check at the middle distal joint (< 0.005f) resets the stick to neutral, allowing rapid "thumb-off" cancellation. The stick vector is normalized ($\|\mathbf{s}\| = 1$) so that axial and diagonal moves share identical magnitude, simplifying scaling of downstream motion.

[1]The thresholds are based entirely on empirical testing during the development by trying different thresholds and subjectively identifying potential issues with them.

Finger Taps. To implement the discrete buttons to trigger actions in the right hand, the index finger is used to trigger Attack and the middle finger the Jump: (1) Attack – thumb–index tip separation $<$ 15 mm (press) | $>$ 20 mm (release), (2) Jump – thumb–middle tip separation $<$ 15 mm | $>$ 20 mm.

4.5 Procedure

After collecting demographic data, participants were introduced to the study. Participants were initially invited to sit down during the session, but were free to stand and move around as needed. Most participants remained seated throughout, though some opted to stand briefly to better explore and interact with the 3D environment. Participants underwent calibration procedures to optimize eye-tracking and hand-tracking functionalities of the Apple Vision Pro. Participants explored a training area with no time limit, practicing movement, jumping, and attacking interactions. Objects similar to those found in the main game were available to facilitate familiarity with controls and gestures. Participants engaged in three consecutive sessions, each involving a distinct input method (two different gesture-based input methods and one controller-based input) and were instructed to (1) Locate and interact with an NPC (non-player character), (2) Complete a given task assigned by the NPC (collect all the bottles and return them to the NPC), (3) Locate and collect a treasure within the virtual environment. The entire gaming session covering all three conditions took approximately 75 minutes, including the training and the gameplay.

4.6 Data Analysis

We used repeated measures ANOVA and t-tests for the parametric data and the Friedman test and Wilcoxon test for post-hoc analysis for non-parametric data with Bonferroni p-value correction.

5 Results

5.1 Task Completion Time

Controller led to shorter task completion times ($Md = 156.78s, IQR = 81.74$) than Knuckle ($Md = 282.26s, IQR = 109.65$) and Palm ($Md = 437.72s, IQR = 263.91$), supported by Friedman test ($\chi^2 = 27, p < .001, \eta^2 = .56$). Post-hoc Wilcoxon signed-rank test showed that all pairwise comparisons were significant ($p < .001$) (Figure 4).

5.2 Number of Jumps & Attacks

Number of jumps was the lowest with Controller ($Md = 51.50, IQR = 30.25$), followed by Knuckle ($Md = 71, IQR = 53$) and Palm ($Md = 81, IQR = 36.75$), supported by a statistically significant Friedman

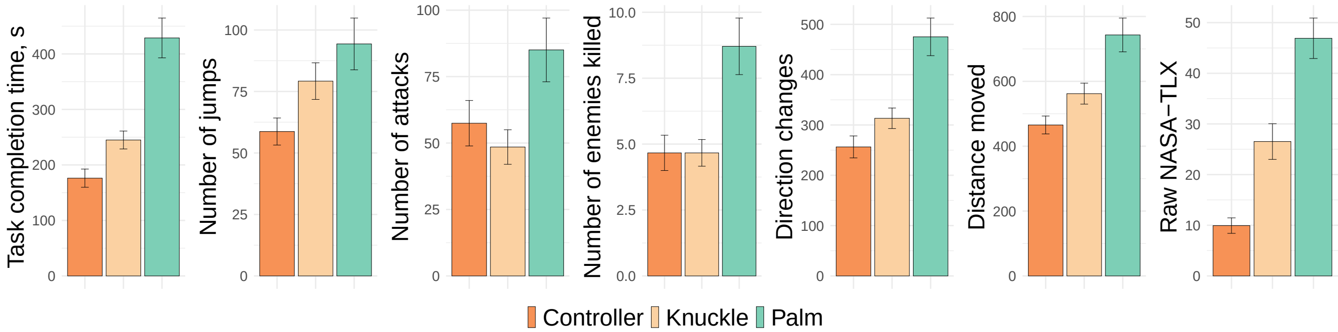

Figure 4: Summary of results for all performance metrics for each input method.

test ($\chi^2 = 15.6, p < .001, \eta^2 = .32$). Post-hoc Wilcoxon signed-rank test showed that all pairwise comparisons were significant: Controller vs. Knuckle ($p < .001$), Controller vs. Palm ($p < .001$), Palm vs. Knuckle ($p < .01$) (Figure 4).

Number of attacks was the highest for Palm ($Md = 67.5, IQR = 63$), followed by Controller ($Md = 44 IQR = 60$) and Knuckle ($Md = 38.5, IQR = 21$), supported by statistically significant Friedman test ($\chi^2 = 9.87, p = .007, \eta^2 = .2$). Post-hoc Wilcoxon test showed that Palm led to significantly more attacks than Knuckle ($p < .001$) and Controller ($p < .001$). No significant difference was found between Controller and Knuckle ($p = .992$) (Figure 4).

5.3 Number of Enemies Killed

To compare enemy-kill counts, Shapiro–Wilk tests indicated normal distributions for Controller ($p = .093$), Knuckle ($p = .223$), and Palm ($p = .201$). A repeated-measures ANOVA was therefore applied. Mauchly's test showed sphericity was violated ($p = .002$), so Greenhouse-Geisser correction was used; the effect of input method remained significant ($F(1.38, 31.7) = 8.08, p = .004, \eta^2 = .2$). Mean numbers of enemies killed were: Palm ($M = 8.52, SD = 5.25$), Knuckle ($M = 4.65, SD = 2.48$), and Controller ($M = 4.67, SD = 3.31$). Paired t-tests revealed that Palm produced significantly more kills than Knuckle ($p = .006$), whereas Palm vs. Controller ($p = .033$) and Knuckle vs. Controller ($p = 1.000$) were not significant at the adjusted $\alpha = .0167$. Thus, only Palm led to a reliably higher enemy-kill count than Knuckle; neither gesture differed from the controller baseline (Figure 4).

5.4 Direction Changes

Direction changes were highest for Palm ($Md = 451, IQR = 207$), followed by Knuckle ($Md = 329, IQR = 138$) and Controller ($Md = 241, IQR = 90$), supported by a statistically significant Friedman test ($\chi 2 = 16.33, p < .001, \eta^2 = .34$). Post-hoc Wilcoxon signed-rank test showed that all pairwise comparisons were significant ($p < .001$) (Figure 4).

5.5 Distance Moved

Participants moved the most using Palm ($Md = 689.31\ units, IQR = 271.68$), followed by Knuckle ($Md = 527.26\ units, IQR = 197.91$) and Controller ($Md = 422.22\ units, IQR = 129.51$), supported by a statistically significant Friedman test ($\chi^2 = 21.33, p < .001, \eta^2 = .44$). Post-hoc Wilcoxon signed-rank test showed that all pairwise comparisons were significant ($p < .01$) (Figure 4).

6 Perceived Workload

Perceived workload was lowest for Controller ($M = 9.93, SD = 7.50$), followed by Knuckle ($M = 26.53, SD = 17.27$) and Palm ($M = 46.91, SD = 19.57$), supported by a statistically significant repeated measures ANOVA ($F(2, 46) = 52.25, p < .001, \eta^{(}2) = 0.49$). Post-hoc tests showed that all pairwise differences were significant ($p < .001$) (Figure 4).

6.1 User Experience Ratings

The summary of Likert scale results is shown in Figure 5.

Intuitiveness. Controller was the most intuitive ($Md = 5, IQR = 0$), followed by Knuckle ($Md = 4, IQR = 0$), and Palm ($Md = 3, IQR = 2$), supported by a statistically significant Friedman test ($\chi^{(}2) = 38.52, p < .001, \eta^2 = .8$). Post-hoc Wilcoxon tests confirmed significant pairwise differences between all methods ($p < .001$).

Ease of Learning. Ratings for how easy the input methods were to learn followed a similar pattern: Controller ($Md = 5, IQR = 0$) > Knuckle ($Md = 4.5, IQR = 1$) > Palm ($Md = 3, IQR = 2$). The Friedman test showed a significant difference ($\chi^{(}2) = 32.90, p < .001, \eta^2 = .62$). Post-hoc comparisons showed all pairwise differences were significant ($p < .001$).

Feeling of Control. When asked if they felt in control of the main character, participants rated the Controller highest ($Md = 5, IQR = 0$), followed by Knuckle ($Md = 4, IQR = 0$), and Palm ($Md = 2, IQR = 1$). The Friedman test was significant ($\chi^{(}2) = 38.28, p < .001, \eta^2 = .79$). All post-hoc comparisons reached significance ($p < .001$).

Perceived Task Performance. Participants' self-assessed performance mirrored previous trends: Controller ($Md = 5, IQR = 0$), Knuckle ($Md = 4, IQR = 0.25$), and Palm ($Md = 3, IQR = 2$). A significant Friedman test result ($\chi^{(}2) = 40.08, p < .001, \eta^2 = .83$) was followed by significant pairwise Wilcoxon results ($p < .001$).

Fun and Engagement. On perceived fun and engagement, Controller scored highest ($Md = 5, IQR = 1$), followed by Knuckle ($Md = 4, IQR = 1$), and Palm ($Md = 3.5, IQR = 1$). The Friedman test was significant ($\chi^{(}2) = 12.42, p = .002, \eta^2 = .26$). However, only the Palm vs. Controller comparison showed a significant difference ($p < .001$).

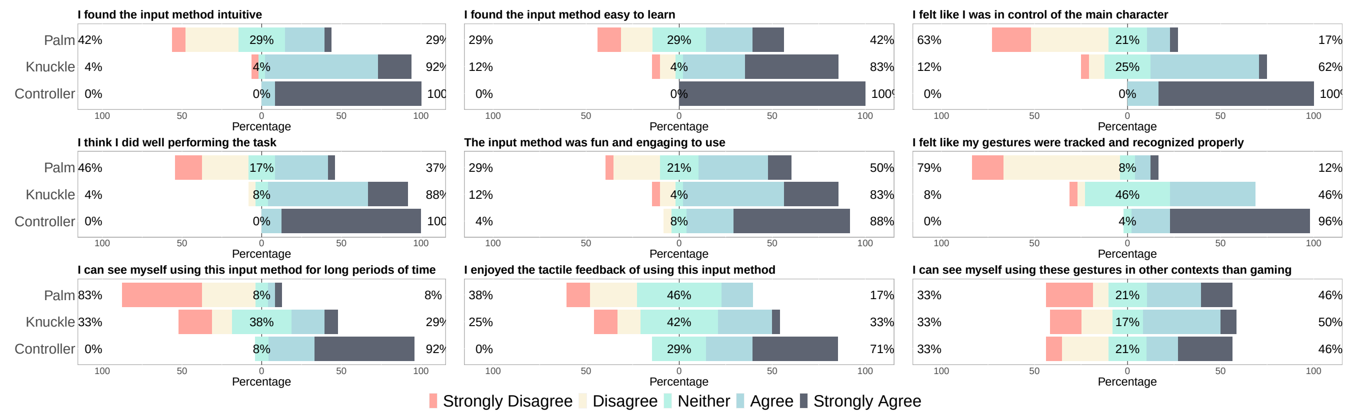

Figure 5: Likert scale results about intuitiveness, ease of learning, feeling of control, task performance, fun and engagement, recognition, long-term use, tactile feedback and public settings.

Tracking and Recognition Accuracy. Participants reported greater confidence in gesture tracking accuracy for the Controller ($Md = 5, IQR = 0.25$), compared to Knuckle ($Md = 3, IQR = 1$) and Palm ($Md = 2, IQR = 0$). A Friedman test showed significant differences ($\chi^{(}2) = 38.84, p < .001, \eta^2 = .8$). All Wilcoxon comparisons were significant ($p < .002$).

Long-Term Usability. Participants were also asked whether they could see themselves using the input method for extended periods. The Controller ($Md = 5, IQR = 1$) was rated significantly higher than both Knuckle ($Md = 3, IQR = 2$) and Palm ($Md = 1.5, IQR = 1$)($\chi^{(}2) = 32.12, p < .001, \eta^2 = .67$). All post-hoc differences were significant ($p < .001$).

Tactile Feedback Enjoyment. When rating the enjoyment of tactile feedback, Controller again led ($Md = 4, IQR = 2$), followed by Knuckle ($Md = 3, IQR = 1.25$) and Palm ($Md = 3, IQR = 1$), supported by a significant Friedman test ($\chi^{(}2) = 23.11, p < .001, \eta^2 = .48$). All post-hoc differences were significant ($p < .001$).

Applicability Beyond Gaming. Participants were asked if they could see themselves using the input methods in non-gaming contexts. Median ratings were: Controller ($Md = 3, IQR = 3$), Knuckle ($Md = 3.5, IQR = 2$), and Palm ($Md = 3, IQR = 2.25$). The Friedman test showed no significant differences ($\chi^{(}2) = 2.45, p = .29$).

6.2 Qualitative Feedback

Overall, participants preferred to play the game using a controller (N = 15), followed by the Knuckle (N = 8) and Palm (N = 1) due to **low effort**. They mentioned that the controller was *"the simplest"* [P8] and one did not have to monitor their hands: *"I don't need to focus on my right hand all the time"* [P10]. Majority (N=16) found Knuckle the most fun and engaging to use, followed by Controller (N = 6) and Palm (N = 2), because it was something unfamiliar: the knuckle input was *"new and fascinating,"* [P3], *"fun to try something new"* [P9], hand movements felt *"easy and intuitive"* [P15] and *"pretty easy to learn"* [P16]. For seven respondents, fun hinged on a **sense of control, precision or speed**: P7 valued the Controller's *"quickness"* but Knuckle was fun *"due [to] the perceived increased accuracy."* [P21]. Three participants enjoyed the **cognitive headroom** afforded by the physical buttons: the Controller *"gave me headroom to take in the game and scene"* [P1].

6.2.1 *Controller.*

22 participants found the Controller the most precise method, followed by Knuckle (N = 2). Most answers highlighted **reliability** and said the buttons *"had no errors"* [P6] and *"never failed"* [P19], whereas gesture recognition sometimes missed or mis-interpreted actions. A second reason was **familiarity**: after *"a lifetime spent with controllers"* [P20] or simply because they were *"used to"* the interface [P12, P14], participants trusted each press to do exactly what they expected. Eight respondents stressed the **ease of control**, describing that the physical controller was *"easier to steer"* [P8] and instantly conveyed a *"feeling of control"* [P7]. Four highlighted the importance of **tactile feedback**, arguing that physical buttons made the tactile feedback clearer.

6.2.2 *Palm.*

21 participants found the Palm method the most fatiguing due to direct **physical discomfort**: users reported *"tension in the hand and fingers"* [P8], sore fingers [P10], or pain in the palm joints [P24] after only a short session. Related issue was awkward hand posture such as keeping the palm flat [P22], moving the thumb in *"unusual patterns"* [P17] or holding the pinky in a cramped position [P3]. Nine participants linked the discomfort to **poor tracking precision**, saying they had to exaggerate or repeat movements to be recognised, which strained the hand: *"I was really straining my hand in order for it to work."* [P18]. Finally, six answers mentioned **frustration or extra mental effort**: getting gestures to register demanded *"more brain power and focus"* [P14] leading to annoyance [P5] and frustration [P23] rather than enjoyment.

6.2.3 *Knuckle.*

14 participants mentioned they can imagine using the Knuckle me-thod on a bus, outdoors, or while walking due to its **portability**: you *"don't need to bring controllers"* [P5, P13] and *"no need for extra physical equipment"* [P16]. Some mentioned that it looked *"most subtle"* [19] or *"very discrete"* [P5]. However, one participant preferring Controller noted that visible controllers are clearer to by-standers than ambiguous hand signals: *"hand gestures can get confusing for other people around me"* [P4]. Seven participants mentioned Knuckle's "natural feel": *"It feels the most*

natural" [P7] and *"the most natural to use in public"* [P22]. Finally, six answers emphasized **ease or comfort** for Knuckles, describing the method as *"easily accessible"* [P12] or noting that not carrying gear was *"comfortable"* [P17].

6.2.4 Movement. For movement, 22 (out of 24) participants preferred using a physical controller and two the Knuckle. They explained their controller preference by **accuracy and responsiveness**: *"It was the most responsive and least strenuous way of doing it"* [P3] and *"It just worked straight away"* [P6]. Participants also noted **familiarity** and **low effort**: *"I am comfortable and used to it"* [P1], "Easiest way to get the right response" [P8]. **Physical comfort** was also mentioned by six participants; while P12 felt the controller *"did not put any strain"*, P4 reported *"my hands were cramping"* when using gestures. Three participants highlighted reduced **cognitive load**, stating they could *"think about the game rather than how to move"* [P1]. Two participants valued **tactile feedback**, saying that controller buttons had *"best tactile feedback"* [P17]. Only the two Knuckle supporters referenced **fun and novelty**: *"It was new, fun and easier than the palm"* [P15].

6.2.5 Actions. For Actions, twelve participants preferred controller buttons for jumping and attacking, while the others chose on-body. Again, **accuracy and reliability** were the most frequently mentioned considerations: *"more accurately performed intended actions rather than accidental"* [P4] and recognition problems with finger taps were *"off-putting"* [P18]. However, the taps reduced confusion between the different actions: *"I didn't mix up jump and attack with the gestures"* [P2]. **Familiarity** played a strong secondary role: *"It is hard to beat what you already know"* [P3], *"Less input problems and the one I'm used to"* [P23]. One user who preferred the Knuckle remarked *"I didn't have to think about which button to press"* [P8]. In contrast, a participant who preferred Controller stated that finger taps *"took some extra brain-power to figure out"* [P12]. Five participants highlighted **fun and novelty** in favor of finger taps, *"I felt more engaged and in general it was more fun!"* [P5] and *"Fun and new method"* [P16]. **Physical-comfort** comments were rare but split: *"less physical strain using a controller"* [P6] over time, *"Switch's [the controller] button is a bit smaller to my liking."* [P20]. Three remarks touched on **cognitive load or flow**: P1 felt that with buttons they could *"relax and just play,"* while P12 experienced extra mental effort with gestures. P19 mentioned **simultaneous-action constraints**, noting the Controller was *"over-kill"* and prevented jumping and attacking at the same time.

7 Discussion

We found that *Controller* yielded the fastest completion times, fewest direction corrections, and lowest mental workload. Although *Knuckle* was chosen as "most fun" and most preferred to use in public settings, *Palm* was rated least enjoyable. Moreover, *Knuckle* did not differ significantly from the *Controller* on the post-condition Likert scale. We discuss these results in the following.

7.1 Gaming Performance

Participants were not instructed to complete the task as quickly as possible. As a result, their approach to task completion time was considerably different. Some naturally adopted a goal-oriented approach and aimed to complete the course quickly (especially when using the Controller). In contrast, others took a more exploratory approach, spending additional time interacting with and testing the different input methods. This difference in play style inflates the task-time variance, and the task completion time should not be seen as an absolute reflection of method capability. Future studies could time-cap the course or add speed incentives to reduce strategy variance. Participants also killed more enemies, traveled farther, and changed direction more often when using the Palm than the Controller. While this could be seen as a sign of more enjoyment (e.g., spending more time, exploring more), it is likely due to difficulties with the input method. Since enemy spawns reset on a timer, a longer course time allows more enemies to spawn and thus gives the user more enemies to kill. Yet the disproportionate hike in distance and heading changes suggests control-precision deficits: frequent overshoot, course-correction, and "drifting" while the Palm method gesture was mis-tracked. This is further supported by the post-test questions, where fun and excitement were not mentioned for the Palm, but where high counts of frustration and irritation could be seen in the answers. Similarly, NASA-TXL scores for Palm were significantly higher than those of Knuckle and Controller.

7.2 Enjoyment and the novelty effect

Although the Controller outperformed the Knuckle gesture in performance, 18 participants rated Knuckle as more enjoyable, aligning with previous findings on gesture-based interaction's hedonic appeal [25, 32]. This is partly due to novelty, which often boosts initial enjoyment through curiosity and freshness, even if usability flaws are present [31]. Open-ended responses reflected this, with 15 participants describing Knuckle as "fun to try something new" or "fascinating". However, literature warns that such early excitement declines over time [40], making longitudinal testing critical to assess sustained engagement. With continued exposure, gesture systems can become more usable as learning curves flatten and system familiarity increases. This was evident in the case of the one experienced Apple Vision Pro user, who preferred the Controller but completed gesture tasks the fastest—suggesting that system familiarity boosts gesture efficiency over time. Conversely, the Controller benefits from decades of exposure, making it instantly intuitive and effortless. This reflects the mere exposure effect: we favor what is familiar and well-mastered [7]. Familiarity also affects how users assign blame: errors with the Controller were often seen as user mistakes, while similar gesture failures were attributed to system flaws. This asymmetry was especially evident with the Palm gesture method, where some participants exaggerated movements when their input was not recognized. This pushed their thumbs outside the detection range, disabling the system and worsening precision. Many of these issues stemmed from a lack of system feedback, causing users to misinterpret how the gestures worked and unintentionally degrade their accuracy. While novelty drives initial enjoyment of gestures, sustained usability depends on experience, feedback, and system transparency. Meanwhile, the Controller benefits from habitual familiarity, which increases ease of use and shields it from user frustration in ways unfamiliar systems currently cannot.

7.3 Invisibility can outweigh flawless control

Participants who preferred the Controller emphasized practical certainty—describing it as the most responsive, consistent, familiar, and mentally effortless. Many noted it allowed them to focus on gameplay rather than input, with familiarity, low mental load, and reliable accuracy forming the decisive trio—consistent with prior findings [18, 28, 30, 32]. In contrast, those who favored the Knuckle (gesture) method cited fun, novelty, convenience, and enjoyment, rarely mentioning functional metrics. Even when acknowledging the Controller's ease or Knuckle's imperfect accuracy, enjoyment often tipped the balance. Some explicitly weighed both options: P13 chose the Controller for precision but would switch to Knuckle with better tracking; P16 preferred Knuckle despite recognizing the Controller as simpler, calling it "a new experience." These responses reveal a clear trade-off: Controllers dominate in functionality, while gestures excel in experiential appeal—echoing prior studies [28, 32, 43]. Participants shift preference when their personal threshold for either precision or enjoyment is crossed.

This split persisted in public-use scenarios (e.g., on buses or outdoors), but the reasoning shifted. Knuckle was favored for being discreet, non-invasive, and requiring no extra hardware. For some, this outweighed precision concerns. However, others switched to the Controller for its reliability in unpredictable settings. P14, for example, found gestures fun at home but preferred buttons in motion. A seasoned Vision Pro user opted for a controller during longer sessions, citing better accuracy. Notably, users who valued stealth also criticized palm gesture precision, highlighting an accuracy–discretion trade-off: in public, users tolerate some imprecision to avoid drawing attention, up to a point. Although both Knuckle and Palm utilize the thumb rest, we assume that Knuckle provides more favorable one-handed interaction conditions by avoiding gross movements and aligning with natural rotational movements of the thumb joint, leading to more comfort [12, 35].

Preferences were context-dependent. In a quiet test room, Controller users framed their choice in instrumental terms—reliability, cognitive ease, and tactile feedback, aligning with established HCI principles [35]. But in mobile social settings, the same participants often prioritized discretion and portability, finding Knuckle attractive for its invisibility and lack of physical gear [26]. Still, eight users stuck with the Controller in public, citing trust in physical buttons amidst motion or ambiguity. Overall, users apply a personal, context-sensitive cost–benefit logic, weighing functionality against experiential or social value, consistent with prior research [30]. Raising the floor of reliability for gesture systems through better feedback or recognition could shift the balance. Conversely, controllers must address their social footprint to remain viable in mobile MR.

8 Limitations and Future Work

This experiment used a small, convenience-based sample of colleagues and local volunteers. This likely influenced results, as participants were highly curious and engaged with the technology, raising concerns about generalizability to broader populations with varied backgrounds, gaming habits, and motor abilities. Participants could spend as much time as they wanted in the training area, but most did so briefly. Given their familiarity with controller conventions, the limited training likely favored the controller condition and may have exaggerated performance differences. The evaluation was limited to a single, puzzle-oriented task focused on locomotion and trigger use. While useful for isolating basic input, it does not reflect how gesture controls perform in other genres—such as fast-paced action, social, or creative games—where timing, cognitive load, and feedback expectations differ. The reference input device was a pair of Nintendo Joy-Cons. Although similar to popular VR controllers in layout, their small buttons and lightweight design were considered fiddly by some, meaning observed performance differences may partly reflect hardware ergonomics rather than input modality alone. The gesture prototype relied on the Apple Vision Pro's outward-facing cameras and a custom classifier for near-hand micro-motions. Occasional tracking dropouts, especially in peripheral poses, added friction—issues that more advanced sensing systems (e.g., depth-sensor gloves) could reduce. Since micro-gestures require fine motor control, users with tremors, arthritis, or motor impairments may find them tiring or unreliable, limiting accessibility without adaptive filters or alternative gesture sets. False/positive rates and latency were not collected during the study since users' experience was the focus of our work, leaving this part out of the scope and for future work. While our implementation relies on Apple Vision Pro and its tracking system, which is relatively high-end and not universally accessible, the underlying concept of using knuckle and palm micro-gestures is not tied to this device. Other XR headsets already provide sensing capabilities that could support our approach. For example, Meta Quest offers camera-based hand tracking that could approximate knuckle and palm positions, while HoloLens exposes skeletal joint data suitable for similar classification. As hand-tracking APIs continue to converge under standards such as OpenXR, we expect our recognition system to become increasingly portable across XR ecosystems, reducing reliance on premium hardware.

Future studies should include longer trials to assess how gesture performance evolves beyond initial learning, and explore a wider range of game genres to evaluate scalability under different interaction demands. Technical improvements like real-time, user-specific calibration could enhance comfort and accuracy. Comparing a broader range of hardware, including modern VR controllers, would help isolate the impact of input modality from device design.

9 Conclusion

Our findings suggest that gesture input is *not* yet a wholesale replacement for controllers in a puzzle-oriented task focused on locomotion and trigger use in a reachable tabletop mixed-reality adventure gameplay. However, it can potentially be a compelling *complement.* Designers can treat micro-gestures as an optional layer, ideal for short, casual, or mobile interactions, while defaulting to a controller when speed, longevity, or high-stakes precision matter the most. A hybrid solution with seamless hand-to-controller hand-off matches what users naturally employ. Micro-gesture input today occupies a sweet spot between raw efficiency and delightful immediacy. It may not dethrone the physical controller just yet. However, when integrated thoughtfully, it expands the interaction palette of mixed-reality gameplay in ways that players already find worth the occasional slip.

References

[1] Alex Adkins, Lorraine Lin, Aline Normoyle, Ryan Canales, Yuting Ye, and Sophie Jörg. 2021. Evaluating Grasping Visualizations and Control Modes in a VR Game. *ACM Trans. Appl. Percept.* 18, 4, Article 19 (Oct. 2021), 14 pages. doi:10.1145/3486582

[2] Apple Inc. 2024. *Apple Vision Pro available in the U.S. on February 2.* https://www.apple.com/newsroom/2024/01/apple-vision-pro-available-in-the-us-on-february-2/

[3] A. Dinesh Babu, T S Balaji, P Kabilamani, S M Nishaant, and Sruthy R. 2024. Wave & Play: Hand Gestures Redefining User Experience. In *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (Tirunelveli, India, 2024-03-11). IEEE, 260–265. doi:10.1109/ICICV62344.2024.00046

[4] Joanna Bergstrom-Lehtovirta, David Coyle, Jarrod Knibbe, and Kasper Hornbæk. 2018. I Really did That: Sense of Agency with Touchpad, Keyboard, and On-skin Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC Canada, 2018-04-21). ACM, 1–8. doi:10.1145/3173574.3173952

[5] Marios Bikos, Yuta Itoh, Gudrun Klinker, and Konstantinos Moustakas. 2015. An Interactive Augmented Reality Chess Game Using Bare-Hand Pinch Gestures. In *2015 International Conference on Cyberworlds (CW)* (Visby, Sweden, 2015-10). IEEE, 355–358. doi:10.1109/CW.2015.15

[6] S. J. Bolanowski, R. T. Verrillo, and Francis McGlone. 2004. Passive, active and intra-active (self) touch. *Behavioural Brain Research* 148, 1-2 (Jan. 2004), 41–45. doi:10.1016/s0166-4328(03)00157-8

[7] R.F. Bornstein and Catherine Craver-Lemley. 2004. Mere exposure effect. *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory* (01 2004), 215–234.

[8] Michael Brown, Aidan Kehoe, Jurek Kirakowski, and Ian Pitt. 2010. *Beyond the Gamepad: HCI and Game Controller Design and Evaluation.* 197–219. doi:10.1007/978-1-84882-963-3_12

[9] Alvin T.S. Chan, Hong Va Leong, and Shui Hong Kong. 2009. Real-time tracking of hand gestures for interactive game design. In *2009 IEEE International Symposium on Industrial Electronics.* 98–103. doi:10.1109/ISIE.2009.5219910

[10] Xiang 'Anthony' Chen, Nicolai Marquardt, Anthony Tang, Sebastian Boring, and Saul Greenberg. 2012. Extending a mobile device's interaction space through body-centric interaction. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* (San Francisco, California, USA) *(MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 151–160. doi:10.1145/2371574.2371599

[11] Wen-Wei Cheng and Liwei Chan. 2024. BodyTouch: Investigating Eye-Free, On-Body and Near-Body Touch Interactions with HMDs. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 152 (jan 2024), 22 pages. doi:10.1145/3631426

[12] Shariff Am Faleel, Michael Gammon, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. 2021. HPUI: Hand Proximate User Interfaces for One-Handed Interactions on Head Mounted Displays. 27, 11 (2021), 4215–4225. doi:10.1109/TVCG.2021.3106493

[13] Martin Feick, Cihan Biyikli, Kiran Gani, Anton Wittig, Anthony Tang, and Antonio Krüger. 2023. VoxelHap: A Toolkit for Constructing Proxies Providing Tactile and Kinesthetic Haptic Feedback in Virtual Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 104, 13 pages. doi:10.1145/3586183.3606722

[14] Rong Gao, Xin Liu, Bohao Xing, Zitong Yu, Bjorn W. Schuller, and Heikki Kälviäinen. 2024. Identity-free Artificial Emotional Intelligence via Micro-Gesture Understanding. doi:10.48550/ARXIV.2405.13206 Version Number: 2.

[15] S. Geetha, Aditya G, Chetan Reddy M, and Nischith G. 2024. Human Interaction in Virtual and Mixed Reality Through Hand Tracking. In *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (Bangalore, India, 2024-07-12). IEEE, 1–6. doi:10.1109/CONECCT62155.2024.10677239

[16] Sean G. Gustafson, Bernhard Rabe, and Patrick M. Baudisch. 2013. Understanding palm-based imaginary interfaces: the role of visual and tactile cues when browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris France, 2013-04-27). ACM, 889–898. doi:10.1145/2470654.2466114

[17] Juho Hamari and Lauri Keronen. 2017. Why do people play games? A meta-analysis. *International Journal of Information Management* 37, 3 (2017), 125–141. doi:10.1016/j.ijinfomgt.2017.01.006

[18] Asim Hameed, Andrew Perkis, and Sebastian Möller. 2021. Evaluating Hand-tracking Interaction for Performing Motor-tasks in VR Learning Environments. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*. 219–224. doi:10.1109/QoMEX51781.2021.9465407

[19] Chris Harrison, Shilpa Ramamurthy, and Scott E. Hudson. 2012. On-body interaction: armed and dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (Kingston, Ontario, Canada) *(TEI '12)*. Association for Computing Machinery, New York, NY, USA, 69–76. doi:10.1145/2148131.2148148

[20] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 453–462. doi:10.1145/1753326.1753394

[21] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. 50, 9 (2006), 904–908. doi:10.1177/154193120605000909

[22] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. doi:10.1145/2556288.2557130

[23] Ian Hamilton. 2004. *Samsung's Android XR Headset Ships in 2025, and We Went Hands On.* https://www.uploadvr.com/samsung-android-xr-headset-ships-in-2025-hands-on/

[24] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3328–3339. doi:10.1145/3025453.3025523

[25] M.-Carmen Juan, Julen Elexpuru, Paulo Dias, Beatriz Sousa Santos, and Paula Amorim. 2003. Immersive virtual reality for upper limb rehabilitation: comparing hand and controller interaction. 27, 2 (2003), 1157–1171. doi:10.1007/s10055-022-00722-7

[26] Katerina Koleva, Martin Hedlund, Gunilla Berndtsson, Martin Ek, Héctor Caltenco, and Andrii Matviienko. 2025. KnuckleBoard: Knuckle-based Input in Augmented Reality while Sitting, Standing, and Walking. 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 366, 8 pages. doi:10.1145/3706599.3720281

[27] Manuel Lehé, Adrian Kühn, Jessica Bersch, Miriam Nora San Leon Wildner, Umme Jobira Ahmad, and Andrii Matviienko. 2024. VRPaperCrafting: Exploring Child-Friendly Paper Crafting in Virtual Environments using Physical Proxies and Hand Gestures. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction* (Uppsala, Sweden) *(NordiCHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 44, 13 pages. doi:10.1145/3679318.3685379

[28] Lorraine Lin, Aline Normoyle, Alexandra Adkins, Yu Sun, Andrew Robb, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. The Effect of Hand Size and Interaction Modality on the Virtual Hand Illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 510–518. doi:10.1109/VR.2019.8797787

[29] Yi Liu, Hongying Meng, Mohammad Rafiq Swash, Yona Falinie A. Gaus, and Rui Qin. 2018. Holoscopic 3D Micro-Gesture Database for Wearable Device Interaction. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 802–807. doi:10.1109/FG.2018.00129

[30] Tiffany Luong, Yi Fei Cheng, Max Möbus, Andreas Fender, and Christian Holz. 2023. Controllers or Bare Hands? A Controlled Evaluation of Input Techniques on Interaction Performance and Exertion in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4633–4643. doi:10.1109/TVCG.2023.3320211

[31] Alexander G. Mirnig, Magdalena Gärtner, Alexander Meschtscherjakov, and Manfred Tscheligi. 2020. Blinded by novelty: a reflection on participant curiosity and novelty in automated vehicle studies based on experiences from the field. In *Proceedings of Mensch Und Computer 2020* (Magdeburg, Germany) *(MuC '20)*. Association for Computing Machinery, New York, NY, USA, 373–381. doi:10.1145/3404983.3405593

[32] Mathias Moehring and Bernd Froehlich. 2011. Effective manipulation of virtual objects within arm's reach. In *2011 IEEE Virtual Reality Conference*. 131–138. doi:10.1109/VR.2011.5759451

[33] Vimal Mollyn and Chris Harrison. 2024. EgoTouch: On-Body Touch Input Using AR/VR Headset Cameras. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 69, 11 pages. doi:10.1145/3654777.3676455

[34] Niels Christian Nilsson, André Zenner, and Adalberto L. Simeone. 2021. Propping Up Virtual Reality With Haptic Proxies. *IEEE Computer Graphics and Applications* 41, 5 (2021), 104–112. doi:10.1109/MCG.2021.3097671

[35] Donald A. Norman. 2008. *The design of everyday things* (first basic paperback, [nachdr.] ed.). Basic Books.

[36] Min Peng, Chongyang Wang, and Tong Chen. 2018. Attention Based Residual Network for Micro-Gesture Recognition. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (Xi'an, 2018-05). IEEE, 790–794. doi:10.1109/FG.2018.00127

[37] Siddharth S. Rautaray and Anupam Agrawal. 2011. Interaction with virtual game through hand gesture recognition. In *2011 International Conference on Multimedia, Signal Processing and Communication Technologies* (Aligarh, India, 2011-12). IEEE, 244–247. doi:10.1109/MSPCT.2011.6150485

[38] Justine Saint-Aubert, Ferran Argelaguet, and Anatole Lécuyer. 2023. Tangible Avatar : Enhancing Presence and Embodiment During Seated Virtual Experiences with a Prop-Based Controller. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (Sydney, Australia, 2023-10-16). IEEE, 572–577. doi:10.1109/ISMAR-Adjunct60411.2023.00121

[39] Ananya Sharma, Geet Kiran Kaur, Lakshay Verma, and Harleen Kaur. 2023. AuroraSync SAHVI: Enabling Virtual Interaction with Hand Tracking and Voice Commands for Enhanced Human-Software Interaction. In *2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech)* (Banur, India, 2023-12-23). IEEE, 128–134. doi:10.1109/ICACCTech61146.2023.00029

[40] Grace Shin, Yuanyuan Feng, Mohammad Hossein Jarrahi, and Nicci Gafinowitz. 2019. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. 2, 1 (2019), 62–72. doi:10.1093/jamiaopen/ooy048

[41] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow Scotland Uk, 2019-05-02). ACM, 1–15. doi:10.1145/3290605.3300767

[42] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Balagopal Raveendranath, Christopher C. Pagano, Andrew C. Robb, Wen-Chieh Lin, and Sabarish V. Babu. 2023. How Virtual Hand Representations Affect the Perceptions of Dynamic Affordances in Virtual Reality. 29, 5 (2023), 2258–2268. doi:10.1109/TVCG.2023.3247041

[43] Jan-Niklas Voigt-Antons, Tanja Kojic, Danish Ali, and Sebastian Moller. 2020. Influence of Hand Tracking as a Way of Interaction in Virtual Reality on User Experience. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)* (Athlone, Ireland, 2020-05). IEEE, 1–4. doi:10.1109/QoMEX48832.2020.9123085

[44] Mariusz P. Wilk, Javier Torres-Sanchez, Salvatore Tedesco, and Brendan O'Flynn. 2018. Wearable Human Computer Interface for Control Within Immersive VAMR Gaming Environments Using Data Glove and Hand Gestures. In *2018 IEEE Games, Entertainment, Media Conference (GEM)* (Galway, 2018-08). IEEE, 1–9. doi:10.1109/GEM.2018.8516521

[45] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. 2011. A Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. In *Human-Computer Interaction – INTERACT 2011*, Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 559–575.

[46] Yiqi Xiao and Renke He. 2020. The intuitive grasp interface: design and evaluation of micro-gestures on the steering wheel for driving scenario. 19, 2 (2020), 433–450. doi:10.1007/s10209-019-00647-0