Comparing Hand and Controller Avatars with Hand Tracking and Controller-Based Interaction

Natalia Ocampo [*] Carleton University
J. Felipe Gonzalez [†] Carleton University
Robert J. Teather [‡] Monash University

ABSTRACT

Previous research suggests that the congruency between common VR input devices – such as controllers or hand tracking – and their visual representations (e.g., hand or controller avatars) influences user experience and performance. However, the specific effects of input-avatar combinations remain underexplored. We study the effects of common input devices (hand tracking and controllers) and visual representations (hand and controller avatars) on performance and perceived success in target acquisition tasks. We included both grasping and pinching gestures across 16 combinations of input, avatar, and target size. Results indicate that hand tracking benefits from any form of visual representation—even when mismatched—achieving up to 5.8% greater accuracy compared to having no avatar, likely due to its reliance on visual feedback in the absence of a physical prop. Controllers were generally preferred and offered faster task completion. However, mismatched avatars had a stronger negative effect with controllers, particularly when the virtual gesture did not align with the physical action, leading to a 5.6% drop in accuracy compared to the matched condition—suggesting that inaccurate feedback can be more disruptive than having no avatar feedback at all.

Keywords: Virtual Reality (VR), Hand Tracking, Handheld Controllers, Input Device, Avatar Representation, Gesture-Based Interaction, Target Acquisition

1 INTRODUCTION

Modern commercial virtual reality (VR) systems such as the Meta Quest [26], HTC Vive [15] and PlayStation VR [35] rely primarily on controllers as input devices to facilitate user interaction with the virtual environment (VE). Controllers provide precise tracking and the immediacy and reliability of button-based selection, making them well-suited to interactions requiring accuracy and control [17, 23, 31]. However, controllers require users to hold a physical object as a proxy for their real hand, despite offering considerably less natural interaction with objects in the environment. In this sense, controllers cause a disconnect between real-world hand movements and virtual actions, impacting user embodiment and presence [41].

Hand tracking, in contrast, is increasingly common as modern headsets [26] now integrate built-in cameras, so external devices such as the Leap Motion [37] are no longer required. Hand tracking reduces the need for additional hardware and supports more natural gestures in VR, eliminating the need for physical controllers [5]. It benefits from greater similarity to real-world interaction by providing a much more direct correspondence between real hand to virtual hand movements, potentially improving user engagement through more intuitive interactions [5]. This is highly desirable in common VR applications such as gaming, education, training, healthcare, therapy, and physical fitness [10, 14, 20].

[*]e-mail:nataliaocampo@cmail.carleton.ca
[†]e-mail: johannavila@cmail.carleton.ca
[‡]e-mail: rob.teather@monash.edu

Although hand tracking enhances the sense of ownership [2] and improves immersion [19, 23], controllers remain the preferred input device for users. Controllers provide faster and more accurate interaction [11, 12, 23, 25], require lower overall workload in general tasks [11, 19], and foster greater user confidence in task execution [17, 29, 39]. A key factor that may contribute to these differences is the visual representation of the input device in VR. VEs typically provide users with a visual reference to represent their input method/device, typically a virtual hand and/or controller avatar [2, 3]. Avatars are known to enhance immersion and embodiment by strengthening the connection between the user's physical and virtual self [12, 18, 33]. However, while maintaining a balance between physical and virtual interaction can improve engagement and task performance [14, 17], mismatched representations can have a negative impact [23].

VR platforms often face a trade-off between input devices (e.g., hand tracking or controllers) and visual representations (e.g., hand or controller avatars). For example, many applications rely on controller input for better accuracy, yet display a hand avatar to enhance embodiment, despite the mismatched representation. Other applications use hand tracking with matching hand avatars to support natural interaction in scenarios where performance is less critical. Many VR games mismatch visual representations entirely, presenting hands or in-game tools with either controller or hand-based input. The extent to which performance and user experience are influenced by the matching or mismatching of input to visual representation remains poorly understood. Existing research has typically focused on a limited number of input modality combinations, visual representations, and gestures [16, 38]. A more comprehensive understanding of how visual representation interacts with different input devices is still needed to guide VR design decisions.

We present a study evaluating user preferences and performance under varying combinations of input device and avatar representation. We aim to address the question: ***how do different input devices and visual representations influence user experience and performance?*** We recruited 48 participants who completed a series of object manipulation tasks in a custom VR game. The task was modeled after the standardized Fitts' law selection paradigm [7], and employed a methodology similar to previous studies [21, 27]. Participants performed tasks using two common input methods—hand tracking and controllers—executing two typical gestures: pinch and grasp. Visual representation conditions varied based on the presence or absence of a hand avatar and a controller avatar.

To our knowledge, this is the first study to systematically evaluate the impact of avatar representation of the input device across both hand tracking and controller input. While previous work has explored related phenomena [16, 38], they often merge input modality and avatar effects, lacking a controlled and systematic comparison—making it impossible to differentiate their individual effects. The main contribution of our paper is a rigorous and systematic analysis of the effects of input devices (hand tracking and controller) across various combinations of avatar representations (hand and controller avatars) on both task performance and user experience.

2 Related Work

2.1 Controllers

Controllers have been the de facto standard VR input device for decades due to their precise tracking, passive haptics, embedded vibrotactile feedback, and long-standing familiarity from consistent use in gaming and other interactive systems [5, 25, 28, 31]. These attributes are why they continue to be used in the most current commercial VR platforms, such as Meta Quest [26] and HTC Vive [15]. As controllers often use motion sensors, infrared tracking, and pressure sensitive buttons to detect user input and translate them into virtual actions, they offer higher accuracy [17], stability, and responsiveness in object manipulation tasks [22, 23, 30]. Moreover, the use of vibrotactile motors enhance user interaction by providing physical sensations that help with gestures like grasping, pressing, or triggering virtual objects [6, 40]. They enable more efficient task completion and greater precision in selection and trajectory-tracing tasks compared to input methods like hand tracking or mouse and keyboard [17].

Despite these benefits, controllers offer limited embodiment and immersion. Since controllers do not support natural mappings between real hand gestures and virtual actions, users are required to heavily rely on button presses. This lowers the sense of presence and agency in VR experiences [41]. Previous studies [22, 30] highlight that controllers may negatively impact body ownership due to the misalignment between a user's real hand and their virtual hand avatar. In addition to impacting immersion and (indirectly) presence, it also increases cognitive load, as controller interactions feel less intuitive compared to natural hand movements. Additionally, traditional controllers have difficulty facilitating natural interactions due to their inability to support five-finger gestures. At least two fingers are used to grip the controller against the palm, leaving the remaining fingers for button interactions. Nevertheless, individuals may require more fingers for controller support or struggle to perform button interactions due to their hand size. Previous studies [4] have also highlighted the challenges individuals with smaller hands encounter compared to those with larger hands.

2.2 Hand Tracking

Hand tracking allows users to interact with VEs using hand gestures. Instead of relying on physical input (e.g., buttons), hand tracking employs computer vision and sensors to detect and translate hand movements into virtual actions [5]. Compared to controllers, hand tracking offers greater realism as users can engage with the virtual world using a 1:1 mapping between their real hand and virtual hand gestures, without the need to learn controller mappings. This improves embodiment as users perceive their virtual hands as an extension of their real hands. Johnson et al. [16] found that participants manipulating virtual objects when using hand tracking reported higher levels of perceived naturalness compared to controllers. Similarly, Argelaguet et al. [3] reported users preferred free-hand interactions for mid-air tasks as it offered a more direct and engaging experience.

However, hand tracking faces several challenges in tracking fidelity and accuracy. Unlike controllers, which provide stable tracking, hand tracking systems are more susceptible to occlusion, latency, and inconsistent gesture recognition, all of which reduce performance in precision tasks [17]. The absence of tactile feedback with hand tracking also reduces the sense of realism. Even simple vibration offered by controllers provides additional tactile feedback missing with hand tracking. Without it, hand tracking systems lack tactile confirmation when interacting with virtual objects, negatively impacting usability, task performance [25], and user preferences compared to other alternatives [16].

2.3 Visual Representation

Visual representations in VR play an important role in maintaining user engagement and improving interaction fidelity. Avatars representing the user's hands or controllers provide a connection between physical actions and virtual responses. The accuracy of the avatar (in terms of appearance, alignment to the real hand, etc.) affect immersion, realism, and user performance [43]. For instance, Lin et al. [22] report that appearance and behavior of virtual hands significantly influence user perception and control. Users exhibited stronger embodiment when virtual hands closely resembled their real hands in shape and motion. Similar effects have been reported in the use of avatars for controllers. Ponton et al. [30] investigated the effects of controller-avatar alignment, showing that maintaining congruency between the physical controller and the virtual counterpart improves performance and embodiment. Other research [14, 30] suggests congruency between physical actions and visual feedback reduces discomfort and improves task performance. Hibbs et al. [14] studied the effects of visual-physical synchronization in a VR cycling experiment. Their findings suggest that accurate alignment between user movements and avatar actions enhances realism, body ownership, and overall performance. Hanashima et al. [13] similarly demonstrated that visuo-motor-tactile synchrony improves embodiment.

Perhaps the most similar study to ours, Venkatakrishnan et al. [38] explored the interaction between input method and visual representation by evaluating three combinations: controller with a controller avatar, controller with a hand avatar, and glove-based hand tracking with a hand avatar, in a task involving moving doors. Similarly, Johnson et al. [16] investigated this relationship in a ball-sorting task using a pinching gesture. Their study included one hand tracking condition (with a hand avatar) and four controller-based conditions with different avatar combinations. While these studies offer valuable early insights into the impact of input and avatar combinations, they are limited in scope. Specifically, they do not systematically evaluate the full range of possible matching and mismatching combinations between input methods and visual representations. In both cases, the input method is tightly coupled with the avatar type, restricting the ability to isolate their individual effects. Additionally, the analysis is constrained to a narrow set of performance metrics and gesture types. Building on this past work, we aim to provide a more comprehensive evaluation by systematically varying both input method and avatar visual representations. We extend prior work by incorporating multiple gestures and a broader set of objective and subjective performance metrics to better understand how these factors influence user experience and task performance.

3 Methdology

Our study explores combinations of input devices across visual representations, with two common interaction gestures (grasping and pinching). We developed a VR game inspired by the Fitts' law [7] reciprocal selection tasks used extensively as a standard [36] to evaluate pointing devices in HCI and VR [21, 27].

3.1 Participants

Via posters at our university and through social media, we recruited a diverse cohort of 48 participants, aged 18 to 30 years (mean age of 20.1 years), 27 self-identified men (56.3%), 18 self-identified women (37.5%), and 3 self-identified non-binary (6.3%). Concerning visual acuity, all but 2 participants had normal (33 participants, 68.6%) to corrected-to-normal (13 participants, 27.1%) vision. Most (42, 87.5%) were right-handed, while 5 (10.4%) were left-handed, and 1 (2.1%) was ambidextrous. VR experience varied widely: 27 participants (56.3%) had limited exposure (only using VR a few times); 15 participants (31.3%) were novice users; and 6 participants (12.5%) had no prior VR experience. Additionally, 36 (75%)

had never used hand tracking, while 11 (22.9%) had limited hand tracking experience. A single participant (2.1%) reported extensive hand tracking experience. Participants had diverse levels of video game experience, self-reported as playing every day (9, or 18.8%), weekly (8, or 16.7%), occasionally on a monthly basis (11, or 22.9%), yearly (10, or 20.8%), and once a year or almost never (5, or 10.4%).

3.2 Apparatus

3.2.1 Hardware

We used a Meta Quest 3 head-mounted display, which features a field of view of 110 degrees horizontally and 96 degrees vertically, a display resolution of 2064 × 2208 pixels per eye, and a Snapdragon XR2 Gen 2 processor for standalone operation. Notably, it also includes an integrated hand tracking system and two controllers that were used as input devices in the experiment.

Depending on the experimental condition and task, participants used either Quest 3's controller or hand tracking capabilities to perform two different interaction gestures: the grasp gesture and the pinch gesture. Each gesture was performed similarly between input devices, subject to their operational differences. The grasp gesture required participants to clench their hand into a fist with hand tracking, or simultaneously press the GRIP, TRIGGER, and A buttons when using the controller. To perform the pinch gesture with hand tracking, participants would bring together the tips of their index finger and thumb or would simultaneously press the TRIGGER and A buttons when using the controller.

3.2.2 Software

We used Unity 2022.3.1f1 and the Meta XR All-in-One SDK to develop the VR software used in the study. The VE was retrieved from the XR All-in-One SDK. It included an open, minimalist virtual space with a few low-detail items placed within the user's vicinity designed to create a focused environment with minimal distractions. The camera was initially positioned at the center of the room, 3 m above the floor, oriented along the positive x-axis in world space. This setup helped prevent height misalignment between the user and the evaluation environment when entering the virtual world. The evaluation environment was a static planting bed[1] placed directly in front of the camera and presented pre-located interactive buttons, tasks, and instructions. See Figure 1

Figure 1: A 3D planting bed as the interactive VE.

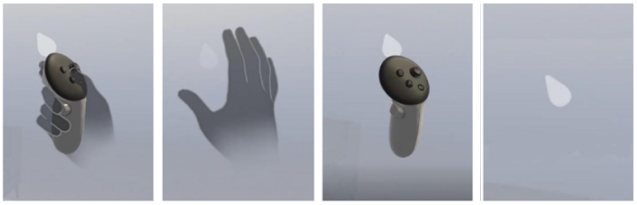
The evaluation environment presented six static virtual rings with a black void within the inner edges to appear as a molehill. Each ring was 0.15 m × 0.07 m × 0.15 m in size, with an outer diameter of 0.15 m and an inner diameter of 0.10 m. The rings were positioned 0.15 m apart from one another, arranged in a circular and equidistant pattern to create a symmetrical, evenly spaced configuration (see Figure 2).

[1]The planting bed model used for the VE was retrieved from Sketchfab at https://skfb.ly/oGGpG.

The setup was in the center of the planting bed angled -8.27° along the x-axis, enabling a high level of comfort for participants arms and necks throughout the tasks.

Figure 2: The six 0.15 m × 0.07 m × 0.15 m rings are placed in a circular and equidistant arrangement for the task.

We varied the distance to and size of the targets, in accordance with the ISO 9241-411 standard reciprocal selection task [1]. According to Fitts' law, varying size and distance yield different task difficulty levels [7, 21, 27], enabling a more accurate representation of everyday tasks participants may encounter, enhancing, enhancing ecological validity. We also incorporated other multi-directional study methods such as Shi et al. [32]. We selected a target distance of 0.25 m and two diameters, 0.15 m (big) and 0.07 m (small). These yielded Fitts' Index of Difficulty (IDs) between approximately 1.4 and 2.1. We intentionally aimed for a low-to-moderate level of difficulty on the ISO standard's recommended range [24] to ensure performance quality and reduce potential confounding effects, as higher difficulty levels are associated with increased error rates and physical strain [32].

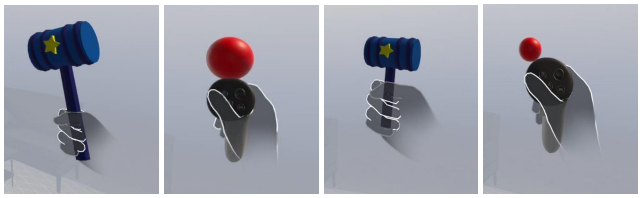


(a) Hand ON and Controller ON (b) Hand ON and Controller OFF (c) Hand OFF and Controller ON (d) Hand OFF and Controller OFF

Figure 3: Visual representations based on combinations of hand and controller avatar.

The software also presented a different visual representation of the hand/controller avatar depending on the condition. The visual representation consisted of all possible combinations of having both the hand and controller avatars being toggled on or off. Thus, the visual representation in a condition could have both hand and controller avatars (Figure 3a), only a hand avatar (Figure 3b), only a controller avatar (Figure 3c), or neither (Figure 3d). Regardless of visual representation, the software always displayed a small cone-shaped cursor (see Figure 3d) to indicate the user's hand/controller position. The software also presented the tasks' targets and interactable objects (a hammer and a ball, used in the two different tasks) as two different sizes, big and small (see Figure 4). The size depended on the condition.

To enhance participant engagement, we added a leaderboard to foster a sense of competition [34]. Competition-based elements,



(a) Hammer (Big) (b) Ball (Big) (c) Hammer (Small) (d) Ball (Small)

Figure 4: The two interactable objects used in the tasks, a hammer and a ball, shown for the big (a, b) and small (c, d) conditions.

such as leaderboards (Figure 1), help motivate participants by challenging their skills, encouraging peak performance [9].

3.3 Procedure

Prior to the main experiment, we conducted a pilot study with 8 participants (distinct from those in the main study) to evaluate the experimental flow and refine the task design. Each participant completed 18 trials of a **pinching** task using a single target size in a simplified virtual environment featuring floating rings. The pilot revealed that participants experienced significant difficulty and discomfort with the pinch gesture, which required high levels of precision and imposed considerable mental and motor effort. Based on these observations, we decided to include a more commonly used —**grasping**—to broaden the scope of our results. Given the challenges observed in the pinch task, we chose to have participants perform the grasp task first in the main study, allowing them to gain familiarity with the environment and reduce potential frustration before attempting the more demanding pinch task.

Upon arrival for the main experiment, participants completed a demographic questionnaire. We then demonstrated the game and explained the experimental tasks (see below). They then completed an informed consent form approved by our university research ethics board. Participants then put on the Meta Quest 3 device and the experiment started. Like the pilot study, participants performed two tasks requiring them to reach towards different rings positioned in a circle. Both tasks always commenced with the participant's hand/controller at the start point marked with an 'X' at the center of the rings. The second location (i.e., the end point) was the target ring highlighted green. Distance between the start and end points was always 25 cm. Participants performed two different tasks covering the most common gestures in object manipulation: the "Whack-a-Mole" task used a grasp gesture, while the "Plug-a-Hole" task used a pinch gesture.

Informed by the pilot, we introduced two different target sizes to incorporate varying difficulty levels and improve the generalizability of the results. The virtual environment was also improved: rings were now embedded into a planting bed with moles (for grasp) and holes (for pinch), and are angled slightly for better ergonomic comfort. We increased the number of trials per condition to 24 to enhance data reliability. Lastly, we refined the success criterion for the pinch task—now, the ball must fully pass through the ring to count as a successful attempt, regardless of whether it touches the ring or not. These modifications aimed to support more natural interactions and generate more meaningful performance data across all conditions.

Grasp Task (Whack-a-Mole) In this task, participants used grasping gestures to pick up a virtual hammer (see Figure 4a/4c). At the beginning of the task, the hammer appeared in arm's reach, and participants grabbed it with the current input technique.

A trial began when participants hit the 'X' mark at the center of the circle of rings using the hammer. This made a ring turn green,

Figure 5: The Whack-a-Mole task; the blue hammer (right) and the red 'X' mark surrounded by the ring setup (center).

indicating it was the target. Then, a 3D model of a mole[2] with a visual indicator appeared at the target location. Participants moved the hammer to the target ring and struck the mole with the hammer to complete the trial. Once the hammer collided with the target mole, the system recorded the distance between the centers of the hammer's face and mole's head. Target ring order followed the standardized Fitts' law reciprocal selection sequence (i.e., always across the circle of targets), starting with the top-right ring.

Pinch Task (Plug-a-Hole) In this scenario, participants used a pinch gesture to pick up and move balls. At the start, a red ball appeared in the center of the rings (i.e., at the start point) and one of the rings turned green, accompanied by a visual indicator, marking the ending point. See Figure 6. Target rings were assigned in the same manner as the Whack-a-Mole task, with the fixed starting point being the left-centered ring. Picking up the ball using the pinch gesture initiated the trial. Participants then moved the ball to the target ring to complete the trial. The trial ended when the center of the ball collided with the target ring, recording the distance between the centers of the ball and ring.

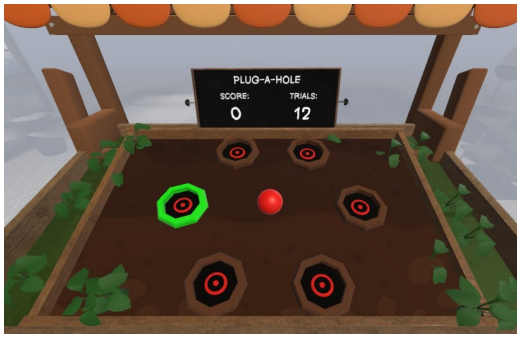


Figure 6: The Plug-a-Hole task; the red ball surrounded by the ring setup (center).

With each new condition, participants first completed 6 non-recorded practice trials, 3 for grasp and 3 for pinch tasks. Immediately following the practice trials were the recorded trials. The recorded trials consisted of 12 trials for each of the Whack-a-Mole (grasp) task, and the Plug-A-Hole (pinch) task.

In total, they completed 24 recorded trials per condition. We included a textual and visual step-by-step guide on how to perform the task and describing the current condition. After completing all trials for a given condition, the experiment would present

[2]The mole model, found within the targets, was retrieved from Sketchfab at https://skfb.ly/ovrVu .

participants with a virtual survey. Participants answered two Likert-scale questions about their experience in the recently completed condition. Participants could take breaks at this time. After completing the survey, the experiment proceeded to the next condition. Participants answered a questionnaire at the end of the experiment related to their overall device, visual representation, sizing, and task preferences. Finally, we compensated them with $15 and thanked them for their participation.

3.4 Design

Participants performed the tasks under different conditions in a $2 \times 2 \times 2 \times 2$ within-subjects design. Independent variables and their levels included:

- **Input**: Controller, Hand Tracking
- **Hand Avatar**: Hand On, Hand Off
- **Controller Avatar**: Controller On, Controller Off
- **Target Size**: Big, Small

The ordering of the 16 combinations of these four factors was counterbalanced according to a balanced Latin square. Given that they completed 24 recorded trials (12 with the grasp task, and 12 with the pinch task) for each of the 16 conditions ($2 \times 2 \times 2 \times 2$ design), participants completed a total of 384 trials (16 conditions $\times$ 24 trials), consisting of 192 grasping trials (Whack-a-Mole) and 192 pinching trials (Plug-a-Hole). The value of total trials excludes the practice trials. The experiment took around 40 minutes on average to complete.

The dependent variables included completion time and accuracy. Completion time (seconds) was the time taken from the start of the task (i.e., touching the red 'X' in the grasp task, or pinching the red ball in the center of the rings in the pinch task) to the end of the task (i.e., hitting the mole in the grasp task, or placing the ball in the hole in the pinch task). Accuracy (cm) was the difference between the center of the hammer's face or ball and the center of the target upon completion of the task. The collected data of our experiment is available at Open Science Framework website[3].

4 Results

We conducted a four-way within-subjects repeated measures ANOVA to analyze completion time and accuracy for both the grasping (Whack-a-Mole) and pinching (Plug-a-Hole) tasks. For all results figures (Figure 7, Figure 8, Figure 9, and Figure 10), error bars show ±1 SE and black horizontal lines (•—•) between bars depict pairwise significant differences via Bonferroni post-hoc tests ($p < .05$). Each bar represents the average performance of the merged big and small target size conditions, grouped by the same input and visual representation combination. This was also applied to Figure 11, Figure 12, and Figure 15. We also analyzed participants' subjective responses using the Aligned Rank Transform (with ANOVA) [42] test with Bonferroni correction for post-hoc testing [8] to compare the different interactive approaches.

4.1 Whack-a-Mole: Grasp Gesture

Completion Time. There were three significant main effects with the Grasp task. First, there was a significant main effect for input on completion time ($F_{1,32} = 55.15, p < .001, \eta_p^2 = .63$), with a mean time of 0.78 s (SD = 0.057 s) with controller, compared to a mean time of 1 s (SD = 0.083 s) with hand tracking, as seen in Figure 7. The main effect for controller avatar also had a significant effect on completion time ($F_{1,32} = 5.29$, $p < .028$, $\eta_p^2 = .14$). Controller avatar *on* yielded faster completion times (mean of 0.87 s, SD = 0.124 s) versus *off* (mean of 0.92 s, SD = 0.139 s). Finally, target size had a significant main effect on completion time ($F_{1,32} = 19.13, p < .001, \eta_p^2 = .38$). Tasks with big targets were faster (mean of 0.84 s, SD = 0.108 s) than those with small targets (mean of 0.94 s, SD = 0.133 s). The main effect for hand avatar was not significant ($F_{1,32} = .06, p = .81, \eta_p^2 = .002$), suggesting its presence or absence had comparatively little impact on participants' performance speed. No interaction effects were found.

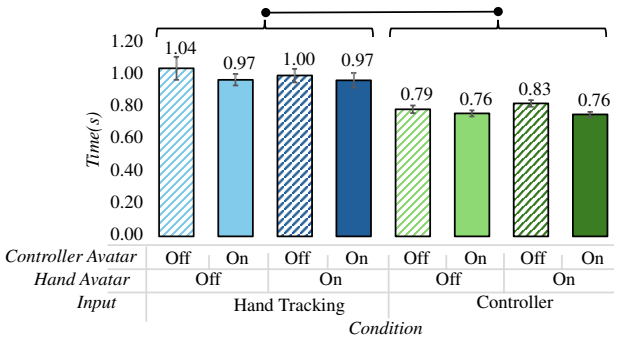


Figure 7: Mean completion time by condition for the Grasp task. Only significant differences between the input is shown for clarity.

Accuracy. ANOVA revealed significant main effects for accuracy for all independent variables, including input ($F_{1,32} = 21.80, p < .001, \eta_p^2 = .41$), hand avatar ($F_{1,32} = 58.82, p < .001, \eta_p^2 = .65$), controller avatar ($F_{1,32} = 6.90, p < .013, \eta_p^2 = .18$), and target size ($F_{1,32} = 201.24, p < .001, \eta_p^2 = .86$). For input, hand tracking (mean of 5.2 cm, SD = 0.19 cm) was more accurate than controller (mean of 5.7 cm, SD = 0.19 cm). With hand avatar *on*, participants were more accurate (mean distance of 5.2 cm, SD = 0.18 cm) compared to not having the hand avatar (mean of 5.6 cm, SD = 0.18 cm). Having controller avatar *off* slightly outperformed having controller avatar *on* (mean of 5.3 cm, SD = 0.17 cm vs. 5.5 cm, SD = 0.16 cm, respectively). Small targets had greater accuracy (mean of 4 cm, SD = 0.03 cm), than large targets (mean of 6.9 cm, SD = 0.07 cm).

There was a significant interaction effect for *hand avatar* $\times$ controller avatar ($F_{1,32} = 26.78, p < .001, \eta_p^2 = .46$), see Figure 8. Pairwise comparisons revealed differences across all combinations except the combinations of controller avatar with and without hand avatar. Visualization with only hand avatar was the most accurate (mean = 5.5 cm, SD = 0.21 cm); having then both controller and hand avatars *on* was next most accurate (mean = 5.5 cm, SD = 0.21 cm). Having both avatars *off* the was least accurate (mean = 5.8 cm, SD = 0.19 cm).

More importantly, there was a three-way interaction effect between *input* $\times$ *hand avatar* $\times$ *controller avatar* ($F_{1,32} = 10.52, p < .003, \eta_p^2 = .25$), as seen in Figure 8. When not using *any* visual avatar representation (i.e., both controller avatar and hand avatar *off*), hand tracking (mean = 5.4 cm, SD = 0.2 cm) offered better accuracy than the controller (mean = 6.2 cm, SD = 0.2 cm). Similarly, when both hand and controller avatars were *on*, hand tracking (mean = 5.2 cm, SD = 0.2 cm) offered better accuracy than the controller (mean = 5.9 cm, SD = 0.2 cm).

Visual representations played a key role with controller input where using no visual representation (i.e., both controller and hand avatars *off*) yielded worse accuracy (mean = 6.2 cm, SD = 0.2 cm) than having only the hand avatar *on* (mean = 4.9 cm, SD = 0.16 cm) or only the controller avatar *on* (mean = 5.7 cm, SD = 0.2 cm). To our surprise, having both avatars (mean = 5.9 cm, SD = 0.2 cm) seemed to have a negative effect compared to having only the hand avatar (mean = 4.9 cm, SD = 0.16 cm).

[3] https://osf.io/urk43/?view_only=19cf112eb7d4440ead996f2719747223

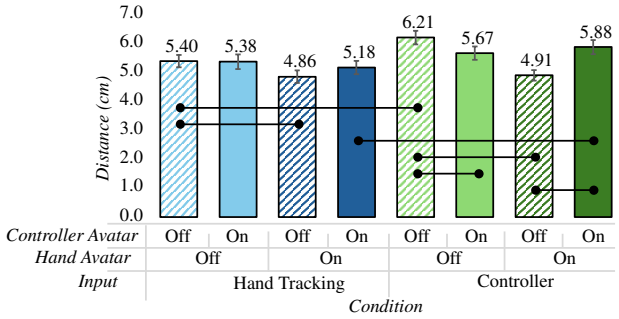


Figure 8: Mean accuracy by condition for the Grasp task.

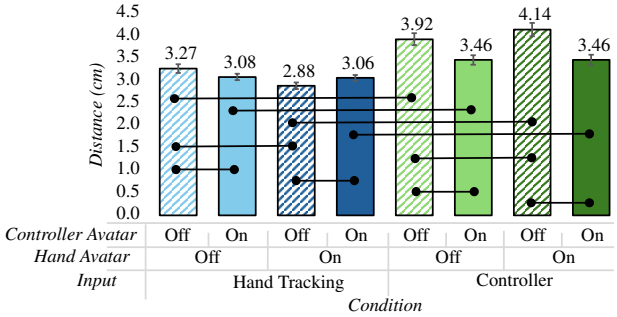


Figure 10: Mean accuracy by condition for the Pinch task.

4.2 Plug-a-Hole: Pinch Gesture

Completion time. For the Pinch task trials (Figure 9), ANOVA revealed relatively few significant effects compared to the Grasp trials. There were significant main effects for input ($F_{1,32} = 79.39, p < .001, \eta_p^2 = .71$) and size ($F_{1,32} = 5.62, p < .024, \eta_p^2 = .15$). Controller input was faster (mean = 0.91 s, SD = 0.037 s) than hand tracking (mean = 1.27 s, SD = 0.05 s). Moreover, big targets yielded a faster time (mean = 1.05 s, SD = 0.047 s) than small targets (mean = 1.14 s, SD = 0.039 s). No interaction effects were found.

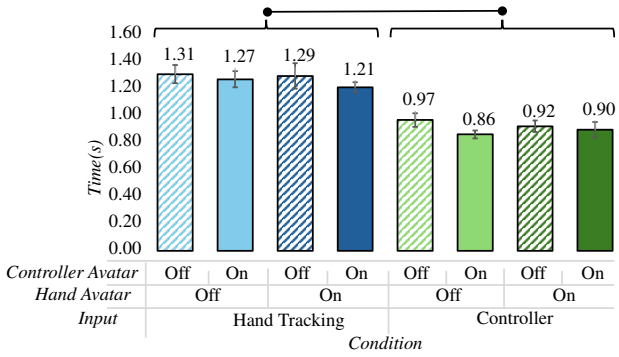


Figure 9: Mean completion time by condition for the Pinch task.

Accuracy. Accuracy results for Pinch trials are seen in Figure 10. ANOVA revealed significant main effects for input ($F_{1,32} = 75.07, p < .001, \eta_p^2 = .70$), controller avatar ($F_{1,32} = 31.41, p < .001, \eta_p^2 = .50$), and target size ($F_{1,32} = 171.90, p < .001, \eta_p^2 = .84$). The main effect for hand avatar was not significant ($F_{1,32} = 1.55, p = .22, \eta_p^2 = .05$) suggesting that it had limited impact on accuracy. Hand tracking input yielded better accuracy (mean of 3.1 cm, SD = 0.05 cm), compared to controller input (mean of 3.7 cm, SD = 0.08 cm). Accuracy was also higher when the controller avatar was *on* (mean = 3.3 cm, SD = 0.06 cm) than *off* (mean = 3.6 cm, SD = 0.09 cm). Similarly, smaller targets offered better accuracy (mean of 2.9 cm, SD = 0.03 cm), than larger targets (mean of 4 cm, SD = 0.06 cm).

The *Input* × *Hand Avatar* interaction effect was statistically significant ($F_{1,32} = 19.46$, $p < .001$, $\eta_p^2 = .38$). Pairwise comparisons revealed differences across all combinations. The most accurate condition was hand tracking with hand avatar *on* (mean = 2.97 cm, SD = 0.06 cm), followed by hand tracking with hand avatar *off* (mean = 3.17 cm, SD = 0.07 cm), then controller input with hand avatar *off* (mean = 3.69 cm, SD = 0.10 cm). The least accurate was controller input with a mismatched hand avatar *on* (mean = 3.80 cm, SD = 0.12 cm). The *Input* × *Controller Avatar* interaction was also statistically significant ($F_{1,32} = 35.37$, $p < .001$, $\eta_p^2 = .31$). Pairwise comparisons showed significant differences across all combinations except between hand tracking and controller avatar *on* and *off*. In this case, hand tracking with controller avatar was slightly more accurate (mean = 3.07 cm, SD = 0.06 cm) than without (mean = 3.07 cm, SD = 0.08 cm), though the difference was not statistically significant. In contrast, controller input with controller avatar *on* offered better accuracy (mean = 3.46 cm, SD = 0.10 cm) than with controller avatar *off* (mean = 4.03 cm, SD = 0.13 cm).

There was a significant three-way interaction between *Input* × *Hand Avatar* × *Controller Avatar* ($F_{1,329} = 15.84$, $p < .001$, $\eta_p^2 = .33$), as seen in Figure 10. Post-hoc comparisons revealed consistent differences between input types across all hand avatar × controller avatar combinations, confirming the main effect of input: hand tracking was more accurate than controller input in every case. Differences between controller avatar conditions were also found across all input × hand avatar combinations, consistent with a main effect, although the direction varied depending on the combination. For instance, controller avatar *on* improved accuracy when using hand tracking with hand avatar *off* (mean = 3.08 cm, SD = 0.07 vs. mean = 3.27 cm, SD = 0.09), controller input with hand avatar *off* (mean = 3.46 cm, SD = 0.12 vs. mean = 3.92 cm, SD = 0.13), and controller input with hand avatar *on* (mean = 3.46 cm, SD = 0.12 vs. mean = 4.14 cm, SD = 0.14). In contrast, when using hand tracking with hand avatar *on*, also having the controller avatar *on* resulted in worse accuracy (mean = 3.06 cm, SD = 0.07) than having the controller avatar *off* (mean = 2.88 cm, SD = 0.07). This suggests that for hand tracking, a matching representation (hand avatar only) leads to better accuracy than a mismatched one (hand + controller avatar), and a mismatched (controller avatar only) is preferable to having no avatar at all. Conversely, for controller input, the presence of the controller avatar improves accuracy regardless of whether the hand avatar is present or not.

Finally, we found significant differences between hand avatar conditions when the controller avatar was *off*: with hand tracking, having only a hand avatar led to better accuracy (mean = 2.88 cm, SD = 0.07) than no avatar (mean = 3.27 cm, SD = 0.09), whereas with controller input, accuracy was higher without any avatar (mean = 3.92 cm, SD = 0.13) than when using only the mismatched hand avatar (mean = 4.14 cm, SD = 0.14). This suggests that with hand tracking, a mismatched representation (only hand avatar) is better than no avatar at all. In contrast, with controller input, it is preferable to have *no avatar at all* rather than a mismatched one.

4.3 In-study Subjective Responses

Upon completing each condition, participants answered the question *"How real/natural did the ∗input type∗ paired with the ∗avatar∗ feel?"*. Participants' responses were ranked on a 5-point Likert scale where 1 meant "Felt Unfamiliar/Artificial" and 5 meant "Felt Like My Real Hand." We analyzed the data using an Aligned Rank Transform (ART) ANOVA [42], revealing significant effects of Input ($F_{1,329} = 76.44$, $p < .001$), Hand Avatar ($F_{1,329} = 4.23$, $p = .040$), and Controller Avatar ($F_{1,329} = 6.61$, $p = .011$). We examined the

proportion of positive responses (scores 4 and 5). Conditions using controller input were perceived as more realistic, with 77.1% of responses being positive, compared to 62.0% for hand tracking. With hand avatar *on*, 71.9% of responses were positive, versus 67.2% when hand avatar was *off*. Finally, for the controller avatar, the proportion of positive responses was similar whether the avatar was shown (69.8%) or not (69.3%). However, the proportion of strongly negative responses (scores 1 and 2) differed: 13.5% when the controller avatar was present, compared to 10.4% when it was absent.

The two-way interaction *input* × *hand avatar* ($F_{1,329} = 10.89$, $p = .001$) was found to be significant. Post-hoc comparisons were conducted using the Aligned Rank Transform Procedure for Multifactor Contrast Tests (ART-C) [8] with Bonferroni correction. All controller input conditions differed significantly from all hand tracking conditions (all $p < .01$). Controller input showed higher perceived realism with hand avatar *on* (78.1% positive responses) and *off* (76.0%), compared to hand tracking, with hand avatar *on* (65.6%) and *off* (58.3%). Similarly, results showed the interaction Input × Controller Avatar ($F_{1,329} = 5.49$, $p = .020$) to be significant. Post-hoc test showed that both controller input conditions differed significantly from both hand tracking conditions (all $p < .001$). Controller input resulted in more positive responses both when the controller avatar was *on* (79.2%) or *off* (75.0%), compared to hand tracking, with controller avatar *on* (60.4%) and *off* (63.5%).

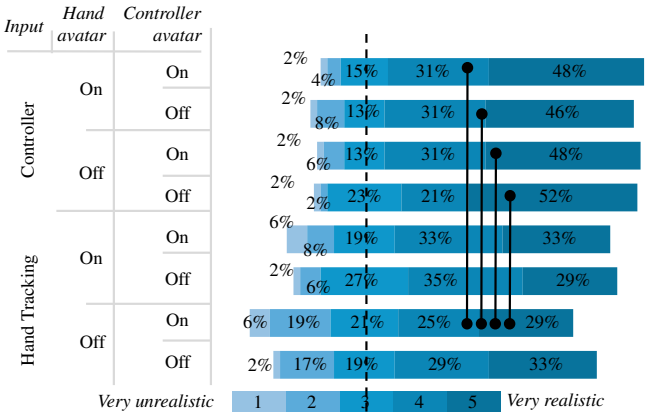


Figure 11: Participants' response on perceived sense of realism for each condition. Black vertical bars show pairwise differences via post-hoc tests ($p < .05$).

A significant three-way interaction was observed for *input* × *hand avatar* × *controller avatar* ($F_{1,329} = 7.14$, $p = .008$). Post-hoc comparisons showed significant differences between all controller combinations and one hand tracking condition ($p < .01$). Controller input had higher perceived realism with both avatars *on* (79.2% positive), with only the hand avatar *on* (77.1%), with only the controller avatar *on* (79.2%), and with both avatars *off* (72.9%) compared to hand tracking using only the mismatched controller avatar (54.2%), which was perceived as the least realistic condition overall (see Figure 11).

Participants also answered *"How quickly did you get used to ∗input type∗ paired with the ∗avatar∗?"* on a 5-point Likert scale, where 1 meant "Took Me A While" and 5 meant "Immediately." An ART ANOVA revealed significant main effects of Input ($F_{1,329} = 12.00$, $p < .001$) and Controller Avatar ($F_{1,329} = 4.65$, $p = .032$). For the Input, controller input led to faster adaptation overall, with 93.8% of responses indicating high familiarity (scores 4–5) compared to hand tracking with 80.2% positive scores and 5.2% indicating perceived slow adaptation. For the Controller Avatar, when it was *off*, 87.5% of responses indicated fast adaptation

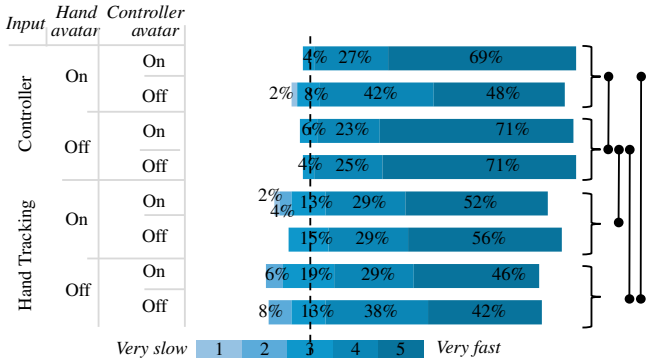


Figure 12: Participants' response on perceived adaptability for each condition. Black vertical bars show pairwise differences between non-different groups designated by parentheses via post-hoc tests ($p < .05$).

compared to 86.5% when it was *on*; in contrast, low adaptation (scores 1–2) was reported in 2.6% of cases without the avatar and 3.1% with it.

A significant two-way interaction was also found for *input* × *hand avatar* ($F_{1,329} = 10.46$, $p = .001$). Post-hoc comparisons (ART-C with Bonferroni correction) revealed significant differences between controller input without hand avatar and all other conditions (all $p < .05$), and between controller with hand avatar and hand tracking without hand avatar ($p = .028$). Controller input without hand avatar had the highest rate of fast adaptation, with 94.8% of responses in the 4–5 range, followed by controller input with hand avatar (92.7%), hand tracking with hand avatar (83.3%), and hand tracking without hand avatar (77.1% fast), as seen in Figure 12.

4.4 Post-Study Survey

At the end of the experiment, participants completed a survey about their experience with the input devices and overall preferences.

Input devices. Participants rated how successful they felt using the input devices on a 5-point Likert scale, where 1 meant "Not Successful" and 5 meant "Very Successful." See Figure 13. A Wilcoxon signed-rank test ($Z = -4.57$, $p < .001$) revealed that participants perceived themselves as significantly more successful with controller input, with 92% of them rating between 4 and 5 compared to the hand tracking, where these scores obtained 69%.

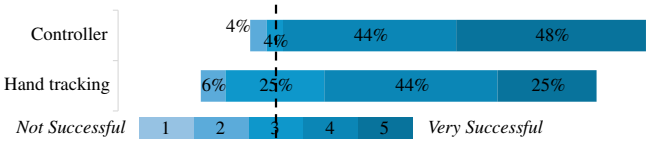


Figure 13: Participants' answers related to the perceived self-performance of the tasks on the different input devices.

Participants also rated the perceived difficulty level when performing the tasks with each input type. See Figure 14. A Wilcoxon signed-rank test ($Z = -4.61$, $p < .001$) revealed that participants perceived tasks as less difficult when using the controller input, with 88% of responses between 1 and 2, compared to 40% for the hand tracking results.

Participants indicated their most and least preferred avatar representations for each input type. See Figure 15. With controller input, half of the participants preferred having both hand and controller avatars, while the least preferred option was having no avatar at all. For hand tracking, 44% (21 participants) selected having both avatars as preferred option, followed closely by 40%

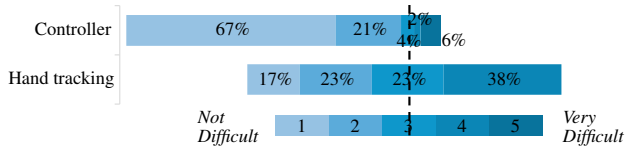


Figure 14: Participants' answers related to the perceived tasks' difficulty on the different input devices.

(19 participants) who preferred only the hand avatar. Similar to controllers, the least preferred option was having no avatar, with 42% (20 participants) expressing dissatisfaction with this condition.

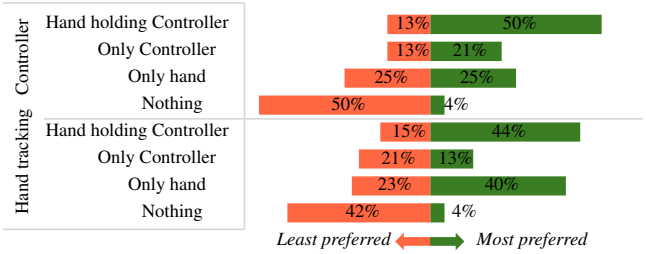


Figure 15: Participants' most and least preferred avatar representation on each input device.

This preference distribution was reflected in participants' open-ended feedback. For controller input, many found the paired hand and controller avatar natural and intuitive. P27 and P7 commented, *"I liked how my hand and controller were represented so it felt very similar to reality... when what I was holding wasn't there it felt like something was missing,"* and *"It felt natural and it looked easy to control and aim accurately,"* respectively. Furthermore, several participants associated the pairing with *"more control"* (P7, P15, P18, P28, P35), *"comfort"* (P10, P22, P23, P33), and *"ease of use"* (P8, P26, P30, P46, P48). While a few participants preferred the controller avatar alone, they noted familiarity and simplicity. For example, P20 noted, *"I felt really natural holding a controller and seeing a controller as my avatar. It just felt right."* In contrast, the no-avatar approach was often described as disorienting. P9 reported, *"I didn't feel like I was using anything... it's like I didn't have much coordination,"* while P7 commented, *"It was hard to keep track of it and it was also difficult to troubleshoot and find out what the problem was."* The hand avatar alone received mixed responses. Some participants appreciated the familiar and realistic visual as it was *"more convenient"* (P41), while others felt it *"made it feel much more like a video game"* (P12) than reality.

A similar pattern emerged with hand tracking. Participants preferred the visual representation of their hands, either alone or holding a controller avatar. P24 explained, *"It just felt the most natural and it mimicked my real hands,"* while P7 said, *"I preferred when it was just my hand, as the motions mirrored my hand's motions identically."* The no-avatar approach was again the most criticized, described as *"confusing"* (P1, P9, P37), *"no sense of navigation"* (P3, P20, P33, P41, P46), and *"less immersive"* (P12, P19, P34). For instance, P27 commented, *"I felt like I was just grabbing nothing."* The controller avatar also received participant dissatisfaction. P21, P6, and P14 questioned its relevance: *"Why would I want a controller visual if I'm using my hand?"*, *"It felt pointless and dumb,"* and *"My hand felt very useless,"* respectively, highlighting the disruption of incongruent visuals on usability.

Participant Preferences. The questionnaire included questions related to preferences regarding certain aspects of the experience. Some notable points include: 35 participants (73%) found the larger target made the task easier, while 13 (27%) preferred smaller targets. A similar trend was observed in task speed, with 39 participants (81%) reporting they were able to complete the task faster with larger targets, whereas 9 (19%) indicated the opposite. When asked about their preferred combination of input device and visual representation, 35 participants (73%) favored using controllers. Among them, 16 preferred both avatars, 8 only the hand avatar, 7 only the controller avatar, and 4 preferred no avatars. Meanwhile, 13 participants (27%) preferred hand tracking, with 5 favoring only the hand avatar, 4 selecting both avatars, 3 preferring only the controller avatar, and 1 opting for no avatar at all. Lastly, the majority of participants (90%) reported enjoying the grasping (Whack-a-Mole) task more than the pinching (Plug-a-Hole) task.

5 Discussion

5.1 User Performance

Grasping tasks were generally faster but less accurate than pinching tasks, although we did not conduct a direct statistical comparison between them. It is possible that the more demanding nature of the pinching task required greater precision, forcing participants to take more time and thereby improving accuracy.

Controller input consistently led to faster task completion than hand tracking in both grasping and pinching tasks. This is consistent with previous findings that physical controllers provide greater stability and motor control due to their tangible structure [10, 16, 18, 24]. The physical buttons and triggers on controllers allow for immediate, discrete input; in contrast, hand tracking relies on continuous motion detection and imprecise gesture recognition [3].

Our findings challenge the assumption that controllers are universally better than hand tracking in terms of both speed and accuracy [13]. While controllers offered faster performance overall, they resulted in significantly lower accuracy, particularly in the pinch task, compared to hand tracking. Participants felt a greater sense of confidence and success when using controllers, reducing concerns about potential errors. When performance is timed and scored, this confidence has led participants to prioritize focus towards speed while heavily trusting the accuracy behind their actions, leading to lower task accuracy. In contrast, hand tracking did not offer such confidence, requiring more intuitive gestures that demanded finer hand adjustments [16]. This complexity may lead to a greater sense of difficulty, increasing the fear of errors. Thus, participants prioritized minimizing mistakes by carefully monitoring and slowing their actions, sacrificing speed.

Avatar representation had a notable effect on input accuracy. In general, matching avatar representations yielded the highest accuracy, outperforming both mismatched and absent avatars. For hand tracking, even a mismatched controller avatar was better than no avatar at all. Interestingly, this pattern differed with controller input and varied depending on the task. With controller input in the grasping task, the best accuracy was achieved with only the hand avatar; in the pinching task, the same condition yielded the worst accuracy. In that case, showing only the controller avatar or both avatars offered better accuracy. These accuracy differences were not reflected in completion time, as timing trends were relatively stable across conditions. This suggests that avatar representation plays a crucial role in controller-based interactions, particularly when there is either a strong match or a noticeable mismatch between the visual and physical representation [38].

One possible explanation lies in the role of haptic feedback and visual alignment with physical actions. In hand tracking, the absence of haptic cues forces users to rely more heavily on visual information—even if mismatched—than on alignment between visual representation and physical sensation. In contrast, controllers offer a tangible object and implicit haptic feedback. When paired with a hand avatar, the visual representation may still feel coherent because it reflects the user's actual hand. However, this may pose limitations depending on the task. Grasping involves a gesture

that naturally aligns with holding a controller, making the hand avatar appropriate in that context. In contrast, pinching requires fine motor control of the thumb and index finger, a gesture that differs significantly from gripping a controller. This likely increases the dissonance between visual and physical input when mismatches occur. Our findings partially support Venkatakrishnan et al. [38], who noted that controller-based hand placement can negatively affect performance. We observed that this influence may be task-dependent, potentially making the performance better or worse. This variation may be explained by how well the physical gesture aligns with the virtual representation. In other words, hand tracking offers more flexibility in gesture design, but its lower reliability increases the need for visual feedback, even if mismatched. Controllers provide more robust input but only benefit from a one-handed avatar representation when the virtual gesture aligns with the physical posture of holding the controller. Otherwise, the mismatch may be counterproductive. Finally, we note that target size had a significant impact on both completion time and accuracy. This is consistent with Fitts' law, where bigger targets lower the target acquisition difficulty and thereby completion time. Small targets yielded higher accuracy than big targets, probably because small targets require greater precision for fine motor control movements [5], and the measurements in bigger targets allowed users to aim farther from the center's target, allowing greater distances.

Our findings suggest that while controllers generally enable faster task performance, this speed can come at the cost of reduced accuracy compared to hand tracking. Avatar representation had a significant impact on accuracy, although no clear effects were observed on completion time. Hand tracking offered greater flexibility in gesture expressiveness, making it a viable option when virtual actions differ from the physical gesture of holding a controller. In contrast, controllers provide robust and fast input, but benefit from careful avatar design. Using only a hand avatar can enhance accuracy when the virtual gesture closely matches the physical interaction; otherwise, it may lead to performance degradation. In such cases, including a controller avatar—or a similarly shaped visual cue—can help align user expectations with physical constraints, potentially improving performance and user experience.

5.2 User Experience

The subjective results revealed a general preference for controllers over hand tracking. Mismatched avatar-input pairings negatively affected perceived realism, as evidenced by the low ratings for hand tracking paired with only the controller avatar. Participants reported similar levels of perceived adaptation speed across conditions; however, the presence of the controller avatar had little impact. Controllers were consistently rated as the easiest to adapt to, particularly when used without the hand avatar. This suggests the pairing between input type and avatar representation influences user perception of embodiment and usability, consistent with prior studies [16, 22, 30, 38].

Participants also expressed a preference towards the grasp tasks over the pinch tasks. We were surprised by how drastic this preference difference was. With hand tracking, this difference may be due to the grasp gesture delivering tactile feedback upon a larger area of the palm and having a lower motor demand as finger extension was not required. In contrast, the pinch gesture requires a precise interaction point between two fingers; grasping uses all fingers in a comparatively less precision-oriented gesture. As for controller input, when performing a grasp, participants knew they needed to interact with all in-contact buttons; this is comparatively easier than recalling the specific buttons required for pinching.

6 Limitations and Future Work

The order of gesture tasks was fixed, with the grasping task always preceding the pinching task. We consciously chose this order based on the anecdotal observation that pinching seemed more difficult in our pilot testing. Although we did not compare gestures directly, this ordering may have introduced learning effects that favored the second (pinching) task, potentially influencing accuracy or user adaptation. Future studies should consider counterbalancing gesture order to control for such effects.

While our subjective questionnaires were designed based on relevant prior work, they were not adapted from validated or standardized instruments. This limits the comparability of our results to other studies using established scales. Future work should incorporate standardized tools for measuring embodiment, usability, and adaptation, such as the Presence Questionnaire or System Usability Scale, to enhance generalizability.

The pinching gesture may have introduced usability challenges. Despite instructions and demonstrations, some participants naturally used pinch variants (e.g., pressing fingers against the palm) that were not detected by the hand tracking system, which required the fingers to be extended. As a result, some pinch attempts were not recognized, potentially affecting both performance and perceived usability. Future research could explore technical limitations in pinch detection methods.

Additionally, a minor software issue was observed by two participants: objects (hammer or ball) could be unintentionally released and lost in the virtual environment if dropped before target contact, slightly increasing completion time. While this bug was reported by only a few participants and likely had minimal impact on overall results, it highlights the need for robust interaction feedback and object constraints in VR task design.

7 Conclusion

Our objective was to enhance understanding of varying visual representations across different input modalities. To this end, we conducted a study evaluating combinations of hand and controller avatars across both hand tracking and controller-based input in grasping and pinching tasks commonly used in VR systems (e.g., games). Although a prior similar study [16] has been conducted, we present a more systematic and thorough exploration of relevant factors by comparing across input type (hand tracking, controller), task (pinch vs. grasp), and visual representation (all combinations of hand avatar on/off and controller avatar on/off). Our evaluation included objective performance measurement (i.e., completion data and accuracy) and subjective perception questions.

Our results confirm previous findings showing faster completion times with controllers compared to hand tracking. However, in terms of accuracy, our findings challenge earlier studies by showing that hand tracking can outperform controllers under certain conditions. Furthermore, although we did not observe a significant interaction between avatar representation and input modality on completion time, we did find a clear effect on accuracy. In particular, matching avatars consistently led to higher accuracy and were more favorably received by participants. Importantly, our results also indicate that the effectiveness of a hand avatar with controllers is task-dependent: accuracy may improve only when the physical gesture aligns well with the visual representation, as seen in grasping tasks, but not in gestures like pinching, where this alignment breaks down.

We note that many VR games (and other non-game VR applications) commonly present avatars that do not match the user's hand or controller. Our findings thus provide valuable insight to future researchers and designers, offering guidance on determining a suitable visual representation for a given input device across different types of common interactions. Overall, our work contributes to ongoing research by exploring and deepening the understanding of the relationship between input devices and avatar representations and their impact on not only each other but user performance and experience as well.

REFERENCES

[1] ISO/TS 9241-411:2012 Ergonomics of human-system interaction Part 411: Evaluation methods for the design of physical input devices, May 2012. https://www.iso.org/standard/54106.html. 3

[2] A. Adkins, L. Lin, A. Normoyle, R. Canales, Y. Ye, and S. Jörg. Evaluating Grasping Visualizations and Control Modes in a VR Game. *ACM Trans. Appl. Percept.*, 18(4):19:1–19:14, Oct. 2021. 1

[3] F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*, pages 3–10, Mar. 2016. ISSN: 2375-5334. 1, 2, 8

[4] M. A. Brown and I. S. MacKenzie. Evaluating Video Game Controller Usability as Related to User Hand Size. In *International Conference on Multimedia and Human Computer Interaction (MHCI 2013)*, Toronto, Ontario, Canada, July 2013. 2

[5] G. Buckingham. Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. *Frontiers in Virtual Reality*, 2, Oct. 2021. Publisher: Frontiers. 1, 2, 9

[6] P.-A. Cabaret, T. Howard, G. Gicquel, C. Pacchierotti, M. Babel, and M. Marchal. Does Multi-Actuator Vibrotactile Feedback Within Tangible Objects Enrich VR Manipulation? *IEEE Transactions on Visualization and Computer Graphics*, 30(8):4767–4779, Aug. 2024. 2

[7] L. D. Clark, A. B. Bhagat, and S. L. Riggs. Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies*, 139:102413, July 2020. 1, 2, 3

[8] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, pages 754–768, New York, NY, USA, Oct. 2021. Association for Computing Machinery. 5, 7

[9] F. S. A. Halim and T. Buhari. The effectiveness of gamification through classroom leaderboards for student engagement, Nov. 2024. 4

[10] A. Hamad and B. Jia. How Virtual Reality Technology Has Changed Our Lives: An Overview of the Current and Potential Applications and Limitations. *International Journal of Environmental Research and Public Health*, 19(18):11278, Sept. 2022. 1, 8

[11] A. Hameed, S. Möller, and A. Perkis. How good are virtual hands? Influences of input modality on motor tasks in virtual reality. *Journal of Environmental Psychology*, 92:102137, Dec. 2023. 1

[12] A. Hameed, A. Perkis, and S. Möller. Evaluating Hand-tracking Interaction for Performing Motor-tasks in VR Learning Environments. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 219–224, June 2021. ISSN: 2472-7814. 1

[13] R. Hanashima and J. Ohyama. How to Elicit Ownership and Agency for an Avatar Presented in the Third-Person Perspective: The Effect of Visuo-Motor and Tactile Feedback. In *Human Interface and the Management of Information: Applications in Complex Technological Environments: Thematic Area, HIMI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings, Part II*, pages 111–130, Berlin, Heidelberg, June 2022. Springer-Verlag. 2, 8

[14] A. Hibbs, G. Tempest, F. Hettinga, and G. Barry. Impact of virtual reality immersion on exercise performance and perceptions in young, middle-aged and older adults. *PLOS ONE*, 19(10):e0307683, Oct. 2024. 1, 2

[15] HTC. VIVE Canada | VIVE XR Elite - Save Big on Base Station-Free PC VR Headset, 2025. 1, 2

[16] C. I. Johnson, N. W. Fraulini, E. K. Peterson, J. Entinger, and D. E. Whitmer. Exploring Hand Tracking and Controller-Based Interactions in a VR Object Manipulation Task. In J. Y. C. Chen, G. Fragomeni, and X. Fang, editors, *HCI International 2023 – Late Breaking Papers*, pages 64–81, Cham, 2023. Springer Nature Switzerland. 1, 2, 8, 9

[17] J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo. Trade-Off between Task Accuracy, Task Completion Time and Naturalness for Direct Object Manipulation in Virtual Reality. *Multimodal Technologies and Interaction*, 6(1):6, Jan. 2022. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute. 1, 2

[18] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoper. Virtual Environ.*, 21(4):373–387, Dec. 2012. 1, 8

[19] A. Kim. A Comparative Study of the User Experience of Controller and Hand-Tracking Interactions in a Virtual Environment. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 744–748, Oct. 2022. ISSN: 2771-1110. 1

[20] H. Kim and Y. Seo. YOGA IN THE METAVERSE: POSSIBILITIES AND LIMITATIONS, May 2024. 1

[21] P. Kourtesis, S. Vizcay, M. Marchal, C. Pacchierotti, and F. Argelaguet. Action-Specific Perception & Performance on a Fitts's Law Task in Virtual Reality: The Role of Haptic Feedback. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3715–3726, Nov. 2022. Conference Name: IEEE Transactions on Visualization and Computer Graphics. 1, 2, 3

[22] L. Lin and S. Jörg. Need a hand? how appearance affects the virtual hand illusion. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '16, pages 69–76, New York, NY, USA, July 2016. Association for Computing Machinery. 2, 9

[23] T. Luong, Y. F. Cheng, M. Möbus, A. Fender, and C. Holz. 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):4633–4643, Nov. 2023. Conference Name: IEEE Transactions on Visualization and Computer Graphics. 1, 2

[24] I. S. MacKenzie. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human–Computer Interaction*, 7(1):91–139, Mar. 1992. Publisher: Taylor & Francis _eprint: https://doi.org/10.1207/s15327051hci0701_3. 3, 8

[25] A. Masurovsky, P. Chojecki, D. Runde, M. Lafci, D. Przewozny, and M. Gaebler. Controller-Free Hand Tracking for Grab-and-Place Tasks in Immersive Virtual Reality: Design Elements and Their Empirical Study. *Multimodal Technologies and Interaction*, 4(4):91, Dec. 2020. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. 1, 2

[26] Meta. Meta Quest MR, VR Headsets & Accessories, 2025. 1, 2

[27] P. Monteiro, H. Coelho, G. Gonçalves, M. Melo, and M. Bessa. Exploring the user experience of hands-free VR interaction methods during a Fitts' task. *Computers & Graphics*, 117:1–12, Dec. 2023. 1, 2, 3

[28] T. Novacek and M. Jirina. Overview of Controllers of User Interface for Virtual Reality. *Presence*, 29:37–90, Dec. 2020. Conference Name: Presence. 2

[29] A. W. Pangestu, C. H. Primasari, T. A. P. Sidhi, Y. P. Wibisono, and D. B. Setyohadi. Comparison Analysis of Usability Using Controllers and Hand Tracking in Virtual Reality Gamelan (Sharon) Based On User Experience. *Journal of Intelligent Software Systems*, 1(2):89–103, Dec. 2022. Number: 2. 1

[30] J. L. Ponton, R. Keshavarz, A. Beacco, and N. Pelechano. Stretch your reach: Studying Self-Avatar and Controller Misalignment in Virtual Reality Interaction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–15, New York, NY, USA, May 2024. Association for Computing Machinery. 2, 9

[31] H.-R. Rantamaa, J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo. Comparison of a VR Stylus with a Controller, Hand Tracking, and a Mouse for Object Manipulation and Medical Marking Tasks in Virtual Reality. *Applied Sciences*, 13(4):2251, Jan. 2023. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. 1, 2

[32] R. Shi, Y. Wei, Y. Li, L. Yu, and H.-N. Liang. Expanding Targets in Virtual Reality Environments: A Fitts' Law Study. 2023. 3

[33] M. Slater, A. Steed, J. McCarthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors*, 40(3):469–477, Sept. 1998. 1

[34] R. Smiderle, S. J. Rigo, L. B. Marques, J. A. Peçanha de Miranda Coelho, and P. A. Jaques. The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments*, 7(1):3, Jan. 2020. 3

[35] Sony. PlayStation VR | Live the game with the PS VR headset, 2025. 1

[36] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing

device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789, Dec. 2004. 2

[37] Ultraleap. Digital worlds that feel human | Ultraleap, 2025. 1

[38] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. How Virtual Hand Representations Affect the Perceptions of Dynamic Affordances in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2258–2268, May 2023. Conference Name: IEEE Transactions on Visualization and Computer Graphics. 1, 2, 8, 9

[39] J.-N. Voigt-Antons, T. Kojic, D. Ali, and S. Möller. 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)*, pages 1–4, May 2020. ISSN: 2472-7814. 1

[40] C. Wee, K. M. Yap, and W. N. Lim. Haptic Interfaces for Virtual Reality: Challenges and Research Directions. *IEEE Access*, 9:112145–112162, 2021. Conference Name: IEEE Access. 2

[41] G. Wilson and M. McGill. Violent Video Games in Virtual Reality: Re-Evaluating the Impact and Rating of Interactive Experiences. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, pages 535–548, New York, NY, USA, Oct. 2018. Association for Computing Machinery. 1, 2

[42] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 143–146, New York, NY, USA, May 2011. Association for Computing Machinery. 5, 6

[43] J. Zhang, M. Huang, R. Yang, Y. Wang, X. Tang, J. Han, and H.-N. Liang. Understanding the effects of hand design on embodiment in virtual reality. *AI EDAM*, 37:e10, Jan. 2023. 2