

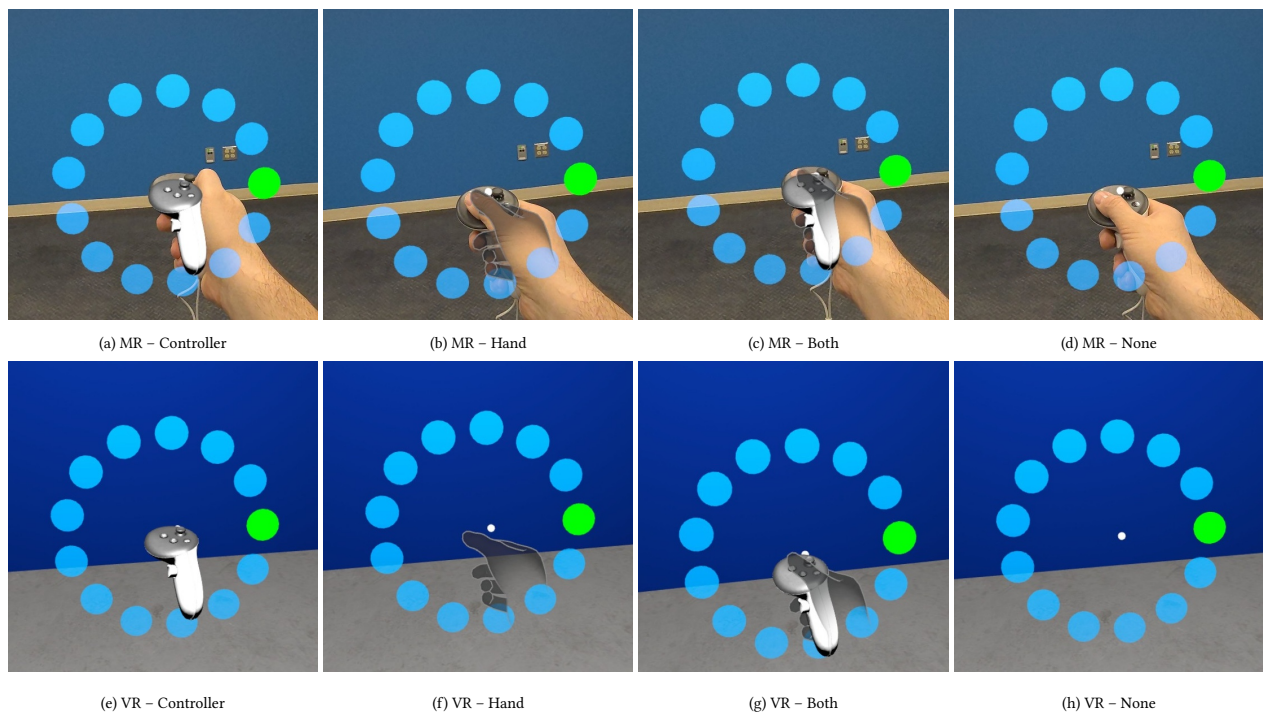
# Effects of Virtual Controller Representation and Virtuality on Selection Performance in Extended Reality

Eric DeMarbre  
Carleton University  
Ottawa, Canada  
eric.demarbre@carleton.ca

J. Felipe Gonzalez  
Carleton University  
Ottawa, Canada  
johannavila@gmail.com

Jay Henderson  
Memorial University  
St. John's, Canada  
jayhend@mun.ca

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



**Figure 1: The 8 (4×2) experimental conditions with controller representation (left to right: controller only, hand only, both controller and hand, and no representation), across reality modes: Mixed Reality (top row) and Virtual Reality (bottom row).**

## ABSTRACT

We present an experiment exploring how the controller's virtual representation impacts target acquisition performance across MR and VR contexts. Participants performed selection tasks comparing four visual configurations: a virtual controller, a virtual hand, both the controller and the hand, and neither representation. We found performance comparable between VR and MR, and

switching between them did not impact the user's ability to perform basic tasks. Controller representations mimicking reality enhanced performance across both modes. However, users perceived performance differently in MR, indicating the need for unique MR design considerations, particularly regarding spatial awareness.

## CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; *User studies*.

## KEYWORDS

Extended Reality, Virtual Reality Mixed Reality, Human-computer interaction

**ACM Reference Format:**

Eric DeMarbre, Jay Henderson, J. Felipe Gonzalez, and Robert J. Teather. 2025. Effects of Virtual Controller Representation and Virtuality on Selection Performance in Extended Reality. 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/10.1145/3756884.3766004>

## 1 INTRODUCTION

Recent extended reality (XR) head-mounted displays (HMDs), such as the Meta Quest [41], integrate cameras that capture the physical environment, enabling video see-through (VST) capabilities that merge real and virtual content [30]. As a result, these devices can support applications across the entire *Reality–Virtuality Continuum* [42, 51]. Moreover, users can dynamically switch between different stages of the spectrum, from augmented reality (AR) scenes, interactive scenes of virtual information in physical worlds with mixed reality (MR), to fully immersive virtual reality (VR) environments [32, 47, 48].

Interactions are central to the user experience in VR and MR, as they enable effective engagement with virtual content and directly impact immersion, usability, and presence [27, 43]. Development frameworks typically offer tools like ray casting, direct touch, or hand gestures, employing hand-tracking or controllers [41]. These tools often display hands or controllers as visual representations to guide interaction. In VR, such representations are central to maintaining immersion and embodiment, especially as they serve as the user's only visual anchor [14, 33, 50]. In MR, the use of avatar-based visual representations has been more limited, likely due to the continued visibility of the user's real hands and controllers. However, some approaches have shown that augmenting the body with virtual elements, such as extended limbs or third-person views, can still enhance embodiment and interaction performance [9, 12, 56]. To date, it remains unclear whether using similar avatar-based visual representations yields comparable user experiences and task performance in MR and VR, given the perceptual and contextual differences between the two environments.

One factor that may influence interaction differences between MR and VR is depth perception. In VR, issues like the conflict between vergence and accommodation and perspective scaling of targets at greater depths can impact pointing performance [18, 29]. In MR, particularly with video see-through, depth perception issues can yield depth cue conflicts between virtual and real elements, leading to higher misjudgment rates and compensatory head movements [61].

Interaction and visual representation function differently in MR compared to VR, as MR users must navigate the spatial alignment of their own perspective, real-world elements, and virtual content, relying heavily on physical cues like depth perception to make sense of these layers [5, 21, 31]. These differences in these mechanisms are not well understood, which poses challenges for applications aiming to support MR-VR experiences with consistent interaction design. For example, previous research has recreated real-world scenarios in VR and enabled transitions between MR-VR modalities [4, 48, 61]. In such cases, replicating the same interaction techniques across platforms without accounting for the perceptual differences

between VR and MR can negatively impact both user performance and experience.

In brief, there is still limited understanding of how different visual representations affect user performance and experience across VR and MR, and to what extent results obtained in one modality can be considered comparable to the other.

To address these gaps, we conducted a study to examine how visual representations affect selection performance in both VR and MR. Participants performed a Fitts' law-based task under two XR modes, VR and MR, and four visual representation conditions: virtual controller, virtual hand, both combined, and no visual feedback.

Motivating our work are the following research questions:

**RQ1:** Are there quantifiable differences in selection time, throughput, and depth error due to differences between virtual and mixed reality?

**RQ2:** Are there quantifiable differences between selection time, throughput, and depth error due to variations in how the controller and hand are represented?

**RQ3:** How do different controller representations influence user subjective preference?

Our principal contribution is a systematic evaluation of the effects of controller representation in target selection performance across VR and MR environments. Our findings inform the fundamentals of selection for developers, designers, and researchers of pass-through mixed reality applications.

## 2 PREVIOUS WORK

### 2.1 Distinguishing Mixed and Virtual Reality

The Reality–Virtuality Continuum refers to a spectrum of displayed environments that combine real and virtual content, with one end presenting a completely real environment and the other presenting a fully virtual environment, which is what we refer to as virtual reality (VR) [42]. Any combination not falling on the extremes is often referred to as *Mixed Reality* (MR) where physical environments are enhanced with interactive digital content. In this paper, we investigate how visual representation affects user experience and performance by comparing interactions in both VR and MR settings.

In this work, we focus on the two sides of the continuum where visual representation may have a significant impact on interaction design: VR, with fully embedded interactive virtual environments, and MR, where digital elements coexist interactively with physical space.

### 2.2 Visual Representation

The use of visual representations may be distinct between VR and MR. In VR, visual representation plays a crucial role in maintaining immersion and improving interaction fidelity [6, 34, 40], especially since VR removes all external visual feedback and the avatar becomes the user's primary reference point [50]. Visual representations—particularly of hands or controllers—help establish a meaningful link between users' physical actions and their virtual counterparts. Research shows that accurate and congruent avatars enhance embodiment, realism, and performance—for example,

matching hand shape and motion boosts embodiment [33]. Similarly, aligning the visual avatar of a controller with its physical form enhances user control and reduces discomfort [16, 49]. Studies also highlight that visuo-motor-tactile synchrony enhances body ownership and realism during interaction [14]. Moreover, matching the visual representation with the input technique can directly affect user performance and experience [20, 57]. These effects are often linked to the user's sense of embodiment and presence [24, 62], as even partial representations—such as virtual controllers—can influence proprioception and the feeling of 'being there' [3, 52].

The uses of avatars in MR are different. Research has explored how avatar representations influence user experience and perception in mixed environments. For example, Genay *et al.* [12] examined third-person perspectives in AR, showing that changes in avatar appearance can affect self-perception, relating to the Proteus effect. Feuchtnner and Müller [9] demonstrated that users can maintain ownership over virtual limbs, even when these are extended or deformed. Similarly, Otono *et al.* [44, 45] found that soft-body transformations and holographic third-person views can enhance the sense of embodiment. In a more task-oriented study, Venkatakrishnan *et al.* [56] showed that interaction performance in dense object spaces improves when users are provided with an augmented self-avatar, highlighting the value of visual hand representations in MR selection tasks.

Most existing studies focus on VR. A recent systematic review found that out of 72 papers on avatar use in XR, only six addressed MR, and just one explored asymmetric VR/MR setups [60]. Moreover, none of these focused on performance in selection tasks. This gap may stem in part from the lack of a consistent definition of MR across studies, as the term encompasses a wide range of technologies—from MR overlays to holographic displays and CAVEs [42]. Some MR-based Fitts' law studies suggest that having a visible body or avatar may support performance [38, 39], but further research is needed to evaluate these effects in more complex MR contexts.

Overall, there remains a lack of direct comparative studies evaluating how different visual representations influence user experience and performance across VR and MR. This limits our understanding of the respective strengths and limitations of visual representations in each environment.

## 2.3 Depth Perception

Depth perception is critical to 3D interaction, but both VR and MR devices have several perceptual problems. The human visual system relies on a combination of monocular and binocular depth cues to estimate spatial relationships in 3D space [17, 59]. However, the accuracy of these cues can be disrupted by the technical limitations of XR technology.

In VR, one of the most significant challenges is the vergence-accommodation conflict, where, unlike in reality, the eyes converge to the depth of a perceived stimulus while accommodating (focusing) to the display surface. This leads to visual discomfort and reduces accuracy in depth-based tasks [8, 18]. Several studies have shown that target depth—the distance from the user to the interaction plane—also influences pointing performance [19, 27, 29].

As targets appear smaller with increasing depth, selection times tend to increase [25].

In MR/AR, depth perception is further complicated by the blending of real and virtual elements. Users must interpret spatial relationships between their egocentric viewpoint, physical objects, and digital overlays [5, 21, 31]. Studies have reported consistent depth underestimation in MR and MR [22, 54], although improvements have been noted in more recent systems [2]. Factors such as mismatched lighting, texture resolution, and contrast between virtual and physical content can mislead depth judgments [31, 55]. Occlusion errors are also a major concern: in MR, virtual content can erroneously occlude physical elements like the user's own hands, disrupting interaction and distorting perceived spatial order [58].

Unfortunately, comparative studies about depth perception between VR and MR are scarce. To our knowledge, no rigorous evaluations have directly contrasted the effects of depth cues on user interaction across VR and MR in equivalent task settings [1]. This limits our understanding of how depth perception issues manifest differently across modalities and how they may affect interaction performance.

## 2.4 Fitts' Law

Since our evaluation employs the ISO 9241-411 standard [11], we describe it and its underlying model, Fitts' law [10], here. Fitts found that the time it takes for humans to acquire targets via rapid aimed movements depends on both the size of the target and the distance to the target from the starting position. While Fitts originally empirically validated this in a one-dimensional task, subsequent reformulations [36] adapted the task to multiple directions in 2D using circular targets arranged in a ring formation. Fitts' law experiments have been effective in comparing baseline performance differences in a variety of conditions, including in XR systems [1] like this work. Figure 2 depicts a standard configuration. Typically, the task involves acquiring a target presented in a different colour from the rest in the ring; the colour indicates which circle the user should select. Movement always proceeds across the ring from one circle to the next, and the diameter of the target ring yields a consistent distance, or amplitude ( $A$ ), for each selection. Target width ( $W$ ) is also varied to produce different task difficulties, presented as the index of difficulty ( $ID$ ) measured in bits, see equation (1):

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

Fitts' law models the relationship between  $ID$  and recorded mean movement time ( $MT$ ) as a linear equation, as seen in equation 2:

$$MT = a + b \times ID \quad (2)$$

where  $a$  and  $b$  are derived via linear regression. Overall selection efficiency is determined using  $ID$  and  $MT$ , yielding throughput ( $TP$ ) measured in bits per second. See equation (3):

$$TP = \frac{ID}{MT} \quad (3)$$

This metric facilitates comparison of the overall effectiveness of an input device or conditions in which it was used to known baselines, supporting the design of new interfaces [35].

Throughput can be further enhanced through the use of so-called “effective” measures, which account for the task participants actually perform in the study rather than that presented. Effective width ( $W_e$ ) replaces the presented target width ( $W$ ) and is calculated through equation (4):

$$W_e = 4.133 \times SD_x \quad (4)$$

$SD_x$  is the standard deviation of selection coordinates along the task axis.  $W_e$  yields an effective target size where 96% (i.e.,  $\pm 2.066$  standard deviations from the mean) of selections would have hit the target, adjusting experimental accuracy and facilitating TP comparison between studies with different error rates. Effective amplitude ( $A_e$ ) is the average of the actual movement distance rather than the presented target distance. Equation (5) includes effective width and amplitude, yielding the effective index of difficulty ( $ID_e$ ), in bits, adjusting for accuracy:

$$ID_e = \log_2 \left( \frac{A_e}{W_e} + 1 \right) \quad (5)$$

Effective throughput,  $TP_e$ , employs  $ID_e$  in the same fashion that standard throughput uses  $ID$ . See equation (6):

$$TP_e = \left( \frac{ID_e}{MT} \right) \quad (6)$$

The main advantage of effective measures is their consistency in throughput calculations. Effective throughput is less susceptible to speed/accuracy tradeoffs [37]. This consistency is important in cross-study comparisons of devices. For example, from extensive prior research employing the standard, the consensus is that the mouse is among the most performant input devices, with a throughput score of around 4.5 bps [15, 26]. Fitts’ law has been used in XR systems to test different input devices and the impacts of variations in VEs on the ability of users to interact with the system effectively [39] and also to examine haptic feedback’s role in VR selection tasks [28].

### 3 METHODOLOGY

In this section, we present the methodology of the user study we conducted using Fitts’ law to evaluate the impact of controller representation in MR.

#### 3.1 Participants

We recruited 40 participants for our study. This included 20 women, 17 men, 1 gender fluid individual, 1 transgender woman, and 1 who declined to answer; the mean age was 23.95 years ( $SD = 6.06$ ). All participants self-reported normal stereoscopic vision and were able to operate one Meta Quest Touch controller. A total of 37 participants reported being right-handed, 2 left-handed, and one reported being ambidextrous but used their right hand for the study. Participants were asked to self-report how frequently they played video games, used virtual reality (VR), and participated in activities that require a high degree of hand-eye coordination (e.g., sports), summarized in Table 1. All participants were compensated with \$15 in local currency for their participation. Our experimental protocol was subject to and passed ethical review by our institution’s ethics board.

Frequency	Video Game	VR	Hand-eye coord. activities
Never	11	5	7
	8	3	13
	14	7	7
	6	7	6
Regularly	1	18	7

**Table 1: Self-reported Likert scale frequency of video gaming, using VR, and participating in activities requiring a high degree of hand-eye coordination. Never = 1, Regularly = 5.**

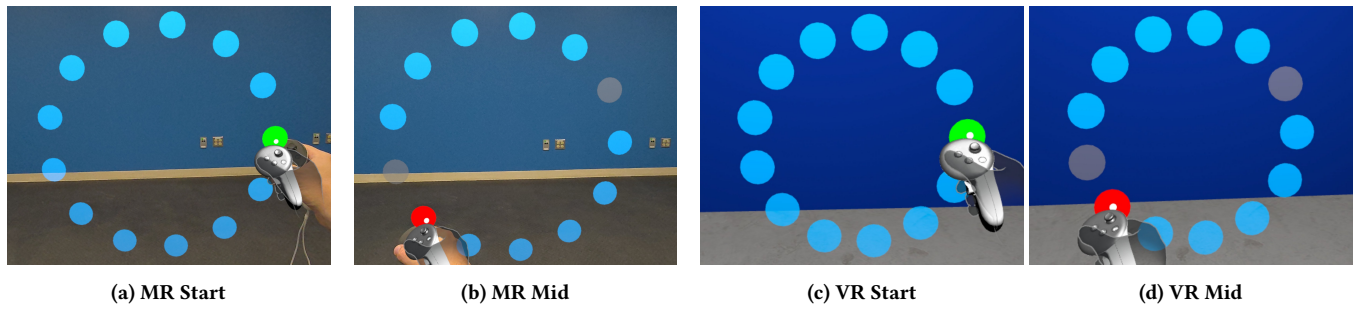
#### 3.2 Apparatus

**3.2.1 Hardware.** We used a Meta Quest 3 VR headset with a 2064 x 2208 pixels per eye resolution, a native 90 Hz refresh rate, and an effective field of view (FOV) of 110° horizontal and 96° vertical. The HMD features an internal processor and storage, allowing it to function without cables, eliminating movement restrictions. The HMD uses external-facing, high-resolution, full-colour cameras with infrared depth sensing for tracking. These cameras also support an MR pass-through mode, which we used for the MR conditions in our experiment. The HMD comes with two tracked hand-held controllers, which we used for the selection task and other interactions in the VEs.

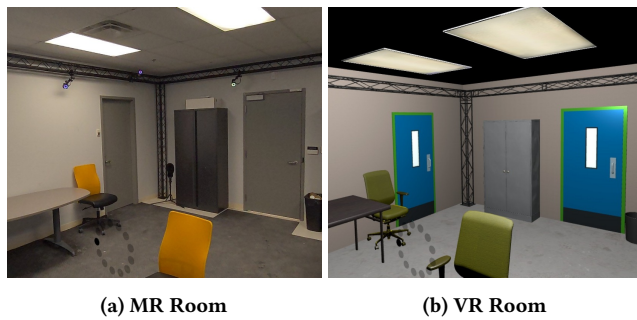
**3.2.2 Software.** We developed the experiment software in Unity 3D (version 6000.0.37f1) using the Meta All-In-One Software Development Toolkit (SDK) (version 72.0.0) for headset and controller tracking, interactions, and access to pass-through cameras. The software presented an extended reality implementation of the standardized ISO-9241-411 task [11]. The application displayed a ring of circular targets in front of participants for selection, as seen in Figure 2. The software presented targets as flat circles to restrict the depth selection plane to the width of the targets. The system supported any combination of target count within a ring, and any combination of amplitudes and widths, as required by the experiment. The system distributed the experiment conditions and the Fitts’ law  $ID$  configurations described in Section 3.4 in balanced Latin squares, selected on the participant number by a modulus calculation. For example, participant number 14 would have condition group 6 ( $14 \bmod 8$ ) and Fitts’ law group 2 ( $14 \bmod 12$ ). The targets changed colour depending on their state: the starting target before the timer started was green (see Figure 2a allowing the participant to rest before beginning if they felt fatigued 2c). Otherwise, the active target to be selected was red, previously selected targets were gray, and the remaining unselected targets were blue (see Figure 2b and 2d). The software also supported having targets optionally turn orange when the poke interactor made contact with them. This feature was enabled only in practice sessions to demonstrate how to make contact with a target. It was disabled during actual experiment sessions, so participants had to judge the depth accuracy of their selections without the extra feedback of the colour change on contact.

Participants selected targets using direct touch via a “poke” interaction. This used the SDK’s default poke interactor object, which is the small sphere visible in Figure 1b. The sphere follows





**Figure 2: The Fitts' Law experiment, at the start and mid points of an ID, in the MR and the VR conditions. The green target indicates the participant can rest to avoid fatigue.**



**Figure 3: The real-world experiment location visible in MR and the VR recreation of the location.**

the tracked controller and provides information on collisions with objects in the virtual environment. This interactor was 1 cm in diameter in the virtual environment, which effectively increases the diameter of the targets by 1 cm. The system used the index trigger on the controller to confirm selection. On selection, if the poke interactor intersected the target, the selection was recorded as a "hit". Otherwise, the selection was considered a "miss" to facilitate the calculation of the error rate. The software played a sound upon pressing the trigger, regardless if the selection was a hit or a miss. Each time the trigger was pressed, the software recorded the selection time, the controller position (captured as a three-dimensional vector), whether the selection hit or missed, as well as the target's amplitude and width. All data was automatically logged in a participant ID and date, time-stamped comma-separated value file. If any unselected targets remained, the next target in the sequence (i.e., the one directly across the ring from the current one, following the ISO standard task) would then activate.

Depending on the condition, the selection task was displayed in either a "VR room" or "MR room", see Figure 3. We referred to this independent variable as XR Mode. The VR environment featured a modestly detailed virtual replica of the room where the study took place, which accurately approximated the exact dimensions and layout of the furniture in the study location (see Figure 3). The second was an MR environment, where the only virtual content displayed was the virtual selection targets and the controller representation. In the MR environment, there were no other visible virtual objects. The Quest 3 pass-through cameras

were activated in this condition, allowing participants to see the real-world room.

We refer to the controller representation independent variable as "Controller Mode". Depending on the Controller Mode condition, the software would change the representation of the tracked controller in the virtual scene, as follows:

**Controller** – The virtual controller representation was a white Meta Quest 3 controller model with a dark grey face (Figures 1a and e).

**Hand** – The virtual hand representation was a dark grey semi-transparent virtual hand model with a light grey outline (Figures 1b and f).

**Both** – Both the virtual controller representation and virtual hand representation were displayed (Figures 1c and g).

**No Representation (None)** – Neither the hand nor the controller were visible. Figures 1d and h).

Note that regardless of the Controller Mode, the Poke Interactor sphere was always visible.

**3.2.3 Experimental Task.** Our experimental task followed the ISO 9241-411 standard [11]. The software was configured to present rings of 13 targets placed at three different amplitudes (distances): 30 cm, 45 cm, and 60 cm, and four widths: 5 cm, 15 cm, 20 cm, and 25 cm. This yielded 12 IDs (see equation 1), providing a range from 1.0 to 3.5, which is within the common range of IDs for XR Fitts' law studies [1], without the amplitude exceeding a comfortable distance for a participant to reach from a seated position. The task involved selecting the highlighted target by reaching out using the controller, touching the spherical cursor at its tip against the active target directly, and pressing the trigger button to select a target. When all 12 IDs for a condition were completed, a UI button appeared, which the participant selected to proceed to the next condition. Participants could take breaks on this next condition screen, or after each set of 13 targets, i.e., before selecting the first target in the sequence, to prevent participant fatigue.

### 3.3 Procedure

We welcomed participants to the experiment space and asked each to complete a consent form and a demographic survey. We then explained the task of the experiment to the participants. Participants remained seated throughout the study. After participants put on

and adjusted the Quest 3, they were presented with a user interface that required them to enter their participant number (to handle condition ordering) and their preferred dominant hand for the entire study. Participants adjusted the position (height and distance) of the targets to reach them all comfortably through a calibration process. Then they completed a practice session for the selection task with 1 block of discs where amplitude = 30 cm and width = 5 cm, and 1 block where amplitude = 60 cm and width = 25 cm. In these practice trials only, the target was highlighted in orange upon intersecting the poke interactor with the target; this feedback was intended to show participants what a successful selection would look like. After completing the practice trials, participants continued to the actual study. Participants were instructed to select the highlighted target as quickly and as accurately as possible. Upon completion of all trials, participants filled out a post-study questionnaire. The questionnaires were custom-designed and asked how *noticeable* the changes in Controller Mode were in VR and MR as a 5-point Likert scale question. This was followed by questions for which Controller Mode was *preferred* in both VR and MR, and which XR Mode was *preferred*. Finally, the participant was asked to provide any additional comments explaining their choices for each of these questions. We created these questionnaires to evaluate participants' subjective perceptions of both the XR and controller conditions, allowing for comparisons with our dependent variables and providing further insights. The participant was then compensated for their participation, and the HMD was then sanitized using a CleanBox and reset for the next participant. The sessions were scheduled for 60 minutes and took, on average, 44 minutes.

### 3.4 Experimental Design

Our experiment employed a  $2 \times 4$  within-subjects design with the factors XR Mode (VR, MR) and Controller Mode (Controller, Hand, Both, and None), as detailed above. We counterbalanced the ordering of the eight combinations of XR Mode and Controller Mode according to an  $8 \times 8$  balanced Latin square. We included 3 target amplitudes (30, 45, 60 cm) and 4 target widths (5, 15, 20, 25 cm) to generate a range of *ID*s per Fitts' law. As we regarded a correct selection as intersecting the 1 cm poke interaction (see Section 3.2.2) with the target, in our Fitts' law analysis and calculation of error rate, widths were 6 cm, 16 cm, 21 cm, and 26 cm [23]. These yielded 12 unique *ID*s ranging from 1.11 to 3.46 bits, distributed using a  $12 \times 12$  Latin square. In summary, participants completed:  $2 \text{ XR Modes} \times 4 \text{ Controller Modes} \times 12 \text{ IDs} \times 13 \text{ Selection Trials} = 49,920$  selections total (or 1,248 selections per participant).

The dependent variables included target selection time (in ms), error rate (% of missed selections), effective throughput (in bps, calculated using Equation 6), and depth deviation (in cm). Depth deviation was calculated as the shortest distance between the poke interactor and the target's *z* coordinate, set to zero when the poke interactor intersects the target. Subjective feedback is reported from the survey described in Section 3.3.

## 4 RESULTS

We conducted data analyses on aggregate data (means) grouped by Controller Mode, XR Mode, *ID<sub>e</sub>* for Throughput (see Equation 6),

and *ID* for Linear Regression Plots (see Equation 1). We conducted Shapiro-Wilk tests of normality on each of our dependent variables and found that our data violated the assumption of normality. Therefore, we performed an aligned-rank transform [63] for each metric, followed by two-way ANOVAs. Post-hoc comparisons were conducted using ART-C [7].

### 4.1 Selection Time

We found a significant main effect of Controller Mode on movement time ( $F_{3,117} = 8.69$ ,  $p < .001$ ,  $\eta_p^2 = 0.18$ ). See Figure 6a. The None condition had the highest movement time ( $M = 0.640$  s,  $SD = 0.300$ ), followed by Controller ( $M = 0.603$  s,  $SD = 0.268$ ), Hand ( $M = 0.600$  s,  $SD = 0.273$ ), and Both ( $M = 0.582$  s,  $SD = 0.238$ ). Post-hoc comparisons revealed significantly higher movement time in the None condition compared to Both ( $p < .001$ ), Controller ( $p = .008$ ), and Hand ( $p < .0001$ ). There were no significant effects of XR Mode ( $F_{1,39} = 0.661$ ,  $p = 0.421$ ), or significant interactions ( $F_{3,117} = 0.624$ ,  $p = 0.601$ ). See Figure 6a. Mean selection time across XR conditions was: MR ( $M = 0.604$  s,  $SD = 0.277$ ) and VR ( $M = 0.608$  s,  $SD = 0.267$ ).

As is commonly done in studies conforming to ISO 9241-9 methodology, we also derived regression models of *MT* in terms of *ID*. The linear regression parameters of these Fitts' law models are seen in Figure 4a, separated by both Controller Mode and XR Mode. These are graphically depicted in Figure 4b (VR mode) and Figure 4c (MR mode).

### 4.2 Effective Throughput

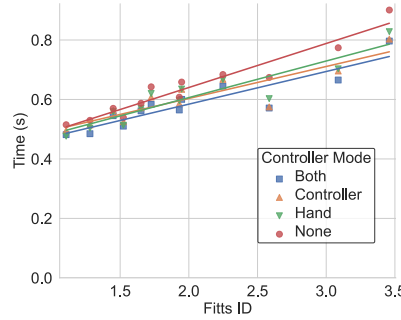
We found a significant main effect of Controller Mode on effective throughput ( $F_{3,117} = 16.33$ ,  $p < .001$ ,  $\eta_p^2 = 0.30$ ). See Figure 6b. The None condition had the lowest effective throughput ( $M = 5.00$  bps,  $SD = 1.55$ ), compared to Controller ( $M = 5.37$  bps,  $SD = 1.60$ ), Hand ( $M = 5.47$  bps,  $SD = 1.67$ ), and Both ( $M = 5.46$  bps,  $SD = 1.55$ ). Post-hoc comparisons revealed significantly lower effective throughput in the None condition compared to Both, Controller, and Hand (all  $p < .0001$ ). There were no significant effects of XR Mode ( $F_{1,39} = 0.341$ ,  $p = 0.563$ ), or significant interactions ( $F_{3,117} = 0.065$ ,  $p = 0.978$ ). See Figure 6b. Effective throughput across XR conditions was: MR ( $M = 5.355$  bps,  $SD = 1.624$ ) and VR ( $M = 5.294$  bps,  $SD = 1.585$ ).

### 4.3 Depth Deviation

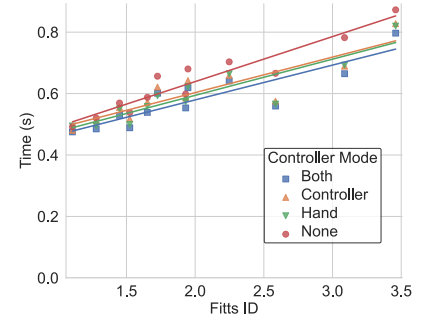
We found a significant main effect of Controller Mode on depth deviation ( $F_{3,117} = 7.24$ ,  $p < .001$ ,  $\eta_p^2 = 0.16$ ). See Figure 6c. The None condition resulted in the highest depth deviation ( $M = 1.41$  cm,  $SD = 1.97$ ), compared to Hand ( $M = 1.25$  cm,  $SD = 2.02$ ), Both ( $M = 1.21$  cm,  $SD = 1.98$ ), and Controller ( $M = 1.20$  cm,  $SD = 1.80$ ). Post-hoc comparisons revealed significantly greater depth deviation in the None condition compared to Both ( $p < .001$ ), Controller ( $p = .005$ ), and Hand ( $p < .001$ ). There were no significant effects of XR Mode ( $F_{1,39} = 2.25$ ,  $p = 0.142$ ), or significant interactions ( $F_{3,117} = 1.91$ ,  $p = 0.132$ ). See Figure 6c. Depth deviation across XR conditions was: MR ( $M = 0.0131$  cm,  $SD = 0.0197$ ) and VR ( $M = 0.0122$  cm,  $SD = 0.0192$ ).

XR	Controller	Intercept	Slope	$R^2$
VR	None	0.344	0.148	0.929
VR	Hand	0.360	0.123	0.849
VR	Controller	0.388	0.107	0.805
VR	Both	0.365	0.110	0.829
MR	None	0.346	0.147	0.905
MR	Hand	0.358	0.118	0.805
MR	Controller	0.372	0.116	0.787
MR	Both	0.353	0.113	0.791

(a) Linear regressions across XR and controller modes.



(b) Linear regression plots in VR by controller modes.



(c) Linear regression plots in MR by controller modes.

Figure 4: Fitts' law analysis by XR mode and controller modes.

#### 4.4 Error Rate

We calculated the error rate as the deviation on the XY plane beyond the edge of the target. Due to a ceiling effect in the error rate data, where most responses were correct, there were insufficient error observations to support a two-factor analysis. To explore potential differences by condition, we conducted separate one-way ANOVAs on the aligned rank-transformed data for XR Mode and Controller Mode individually. There were no significant differences for either XR Mode ( $F_{1,39} = 0.019$ ,  $p = 0.891$ ) or Controller Mode ( $F_{3,117} = 2.47$ ,  $p = 0.065$ ). See Figure 5. For XR Mode, mean error rates were: MR ( $M = 3.57\%$ ,  $SD = 9.08\%$ ) and VR ( $M = 3.64\%$ ,  $SD = 9.18\%$ ). Mean error rates by Controller Mode were: Both ( $M = 3.62\%$ ,  $SD = 9.01\%$ ), Controller ( $M = 3.44\%$ ,  $SD = 9.02\%$ ), Hand ( $M = 3.14\%$ ,  $SD = 8.05\%$ ), and None ( $M = 4.21\%$ ,  $SD = 10.29\%$ ).

#### 4.5 Subjective Results

**4.5.1 Noticeability of changes in controller modes.** From our Noticeability Likert scale metric, an aligned-rank transformed one-way ANOVA revealed changes in controller modes in VR were significantly more noticeable than those in MR ( $F_{1,39} = 6.01$ ,  $p = .019$ ,  $\eta^2 = .13$ ). see Figure 7.

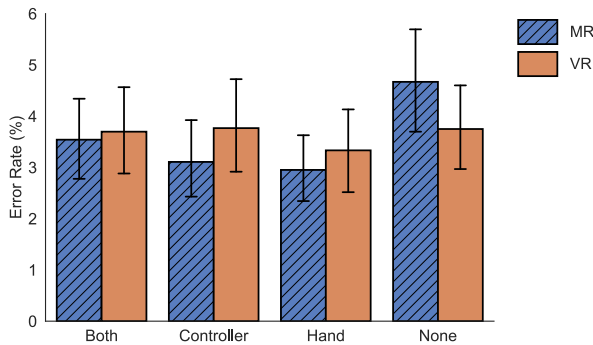
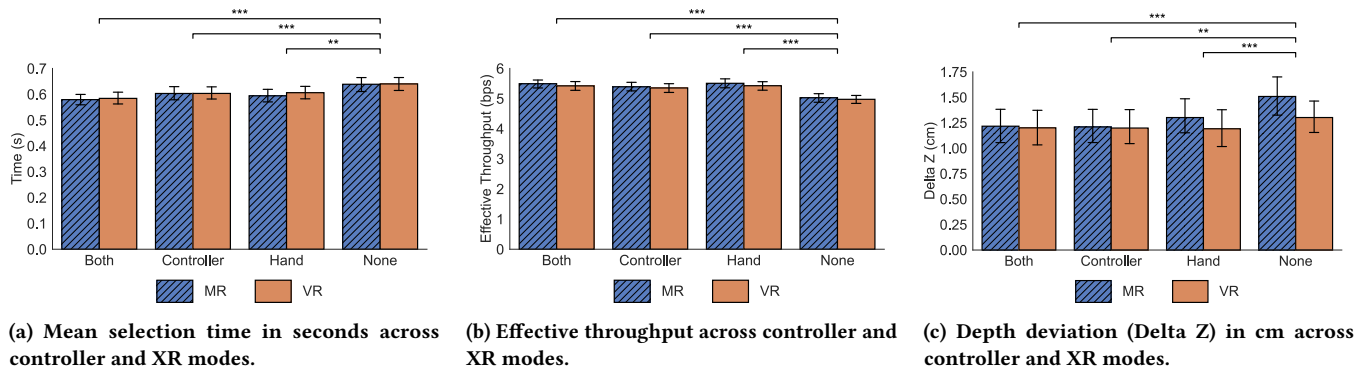


Figure 5: Error rate across controller and XR modes. Error bars indicate a 95% confidence interval.

**4.5.2 Preference and Insights.** Participants completed a post-study preference questionnaire to gather subjective opinions on the conditions. In terms of XR Mode, of the 40 participants, 18 preferred VR conditions, 14 preferred MR conditions, and 8 had no preference. Most participants who preferred the VR mode noted that they felt it was less distracting, as the real-world lab contained additional objects and details that were not visible in the VR mode. Some indicated that they enjoyed the novelty of having the entire environment replaced, rather than seeing the real world. Several participants also noted a visible "lag" or "delay" in the MR conditions, with the controller representations moving slightly behind their physical hand and controller. P27 stated, "The image of the hand and the controller didn't match the real world hand and controller exactly, there was a small amount of lag, and I didn't like seeing a double of them." Participants who preferred the MR mode indicated that they felt more comfortable knowing what was happening around them, compared to the VR mode, where they could not see what was happening in the room. They also reported that seeing their arms made the selection process more realistic and easier, as P36 indicated, "I was more used to those specific conditions where I could gauge things around me better (such as seeing my arms)."

Considering controller mode preference for VR, 18 participants preferred both hand and controller, and 15 preferred only the controller (Figure 8). In the case of both hand and controller, participants indicated that it gave them superior information to assess the position and depth of the poke interactor and also allowed them to judge the angle of rotation of their hand easily. P4 stated, "Having both the hand and the controller as representation made it very easy to keep track of where the controller was at all times." These participants reported that the hand alone seemed unnatural and that it was difficult to judge the position of the poke interactor sphere alone. Those who preferred no representation felt the lack of visual information helped them concentrate on the position of the poke interactor more and felt more accurate (Figure 8). As P17 said, "The hands were distracting since they were a bit weird looking, being not my hand, and when it was only the dot, I had only the dot to see as exact reference instead of a controller as well." Participants who preferred the hand alone cited that its transparency and ability to mimic their hand position made it more realistic.



**Figure 6: Selection Time, effective throughput, and depth deviation (Delta Z) across controller and XR mode. Error bars indicate a 95% confidence interval. Significance is indicated at each level:  $< 0.001 = ***$ ,  $< 0.01 = **$ ,  $< 0.05 = *$ .**

When considering controller representation preference for MR, participants reported an almost equal preference for both hand and controller representation, with 15 selecting it as a preference, and for no controller representation, which received 14 selections. Participants reported that they found the overlapped representation of either the controller or the hand was off-putting, as there was a mismatch with the real-world visual of their hand on the controller. P12 noted, "(The) virtual hand overlaid on top of the real hand felt

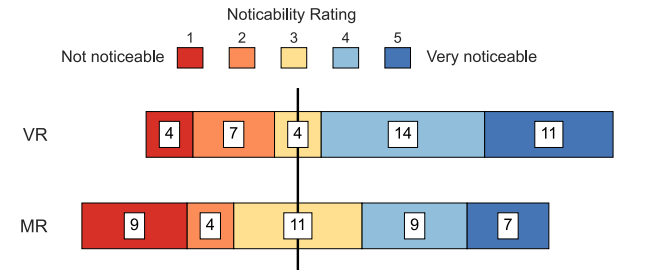
weird because it was duplicated, (I) liked nothing (with just the dot) because you can see your hand and controller." Some participants also noted visible lag in the system, as mentioned earlier, for both conditions and felt that it impacted their performance (Figure 8). Notably, participants who preferred the combination of controller and hand, or "none" for the representation, did not report the lag. Participants who preferred no representation claimed they liked the reduced impact of the lagging visual, which showed just the tracking dot.

## 5 DISCUSSION

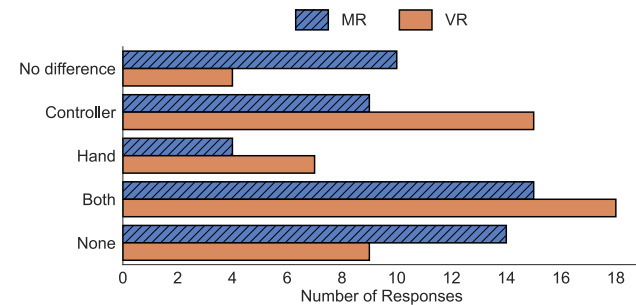
### 5.1 Mixed vs. Virtual Reality

Our first research question asked whether XR mode (VR vs. MR) influences selection task performance. Perhaps the most surprising result was that all measured dependent variables —selection time, effective throughput, depth deviation, and error rate—showed no significant differences between VR and MR modes. While a lack of statistical significance does not indicate equivalence, the means and standard deviations across metrics were also notably similar between MR and VR. This trend is further supported by our Fitts' law analysis (see Figure 6). The linear regression results across VR and MR for all controller modes were very closely aligned, with the largest deviations observed when the virtual controller was present (intercept deviating by only 0.016 and slope by 0.009), which are still considered low. All models showed strong fits to the data ( $R^2 > 0.791$ ). In addition, the visual plots in Figure 4b and Figure 4c exhibit very similar behaviour, with no controller representation (i.e., the None condition) resulting in a noticeably steeper slope. These results suggest that despite concerns about perceptual disruptions in MR, such as latency, occlusion, and other mismatched visuals [5, 31], these issues did not appear to affect selection performance in our study significantly.

Despite these findings, our subjective results indicated that the change to the XR mode may impact a user's *perceived* performance. Participants found changes to the controller representations less noticeable in the MR mode (Figure 7) and reported visible effects of camera or tracking latency in the system. Although neither of these factors appeared to impact performance, they are worth noting, as user testing may reveal performance issues with the system that



**Figure 7: Noticeability ratings of changes in controller modes across XR modes; 1 = Not noticeable, 5 = Very noticeable. Bar labels indicate the number of participants who selected each rating.**



**Figure 8: Number of participants who preferred each controller mode per XR mode (MR or VR).**



users imagine, despite evidence that those performance issues do not exist.

## 5.2 Controller Visualizations

Our second research question asked whether the controller representation had any impact on selection performance. As expected, in comparing no controller representation (the None condition) to all other controller representations, selection time was higher (Figure 6a) and effective throughput (Figure 6b) was lower. Depth deviation (Figure 6c) was also higher with the None condition than all others, suggesting participants had greater difficulty correctly positioning the cursor at the correct depth without the visual context provided by the controller model and / or hand surrogate.

Our Fitts' law regression models (Figure 6) further support these results: for both VR and MR, the None condition regression lines had a steeper slope than all other representations (both, controller, and hand), all of which were largely parallel. This suggests that for more difficult selection tasks (i.e., higher *ID*), selection time increases faster without a controller representation than with any representation is present. These findings are consistent with prior work, where representations like avatars have facilitated performance gains in selection tasks [46, 53].

Our subjective results revealed some nuances when considering controller visualizations. While participants preferred the combined hand and controller representation in both VR and MR, a large portion of participants also preferred no controller representation in the MR mode (Figure 8) despite the performance degradation exhibited in our results. Many of the subjective comments reported in Section 4.5.2 indicated that participants felt the controller and hand-only modes were unnatural. At the same time, the "none" condition followed the movements of their real hand grasping the controller. We would have expected to see a performance impact in MR specifically for these reasons based on previous research [46, 49]. However, the results were similar to VR, suggesting that users may have different criteria for assessing personal performance in MR compared to VR.

## 5.3 Depth Deviation

While noted earlier in Section 5.1, depth deviation did not significantly differ due to XR mode, the differences between MR and VR in the Hand and None controller conditions are notable and warrant further investigation into the components of depth. The lack of visual cues assisting participants in determining depth did have some impact, though not as intense as similar results from previous studies [22, 54]. Improvements in technology, particularly the visual fidelity of the pass-through camera feeds in recent hardware, may be contributing. It is also worth noting that the mean depth deviation in all conditions is high, with a mean of 1.27 cm. With an average target distance of 50 cm, this represents an error size of 0.03. While an approximately 1.25 cm depth deviation may seem minimal, it may still significantly impact real-world applications designed for free movement in a real environment. This is especially important today as MR modes become prevalent on VR devices, considering the relative simplicity of MR application development, particularly on modern Quest platforms. When designing for MR, developers must

consider how this depth deviation may impact the safety of end-users and plan accordingly to provide feedback (visual, auditory, or haptic) to mitigate the effects of depth deviation in the real world.

## 5.4 Limitations and Future Work

To understand the study's implications, we must acknowledge the limitations encountered. Based on the number of participants who reported seeing no changes to controller representation (see Figure 7), the changes may have been too subtle between the conditions besides the no-controller condition. The changes in the XR environment were also not overly distinctive, with an almost exact recreation of the lab where the experiment took place (Figure 3). Comparing more extreme changes in both representations may provide different results. Also, using standard embodiment measurements [13] would allow for clearer assessments of the impressions participants may have on the controller representation in both XR modes.

Secondly, many users indicated experiencing the effects of latency in the MR conditions, where the virtual controller presentation lagged behind the video pass-through image of their real-life hand and controller movement. The specific impact of this latency was not considered. Although it did not appear to have a direct effect on any of the examined performance metrics, its repeated mention by several participants in the subjective feedback suggests that a thorough exploration of the overlapping latencies (tracking and video) may be insightful.

In the interest of building a framework for HCI in MR systems, this research can be extended in several directions. First, future work should measure the impact of conflicting latencies in pass-through-based MR devices and explore methods for compensating for them. Another direction is to investigate variations in controller representation, including differences in model fidelity, size, and realism, to better understand their influence on user perception and performance. The selection task itself can also be expanded upon by incorporating selection confirmation, changing the visual properties of targets, or exploring entirely different tasks altogether.

## 6 CONCLUSION

Our results suggest that selection performance in VR and MR may be fundamentally comparable. Converting from a VR environment to an MR environment does not appear to affect a user's ability to complete basic selection tasks. Similarly, the effects of controller representation on selection performance were consistent across XR modes; this is consistent with prior research and suggests that representations more closely resembling reality are associated with improved selection performance.

However, our subjective results indicate that user *perception* of their selection performance differed between MR and VR, despite the lack of quantifiable differences in performance metrics. This suggests that selection interfaces for MR systems may require distinct design considerations from VR, particularly in terms of spatial awareness. Given the ease with which modern head-mounted displays and SDKs support the porting of VR applications to MR, this is a particularly timely area for future work.



- [37] I. Scott MacKenzie and Poika Isokoski. 2008. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (CHI '08). Association for Computing Machinery, New York, NY, USA, 1633–1636. <https://doi.org/10.1145/1357054.1357308>
- [38] Andrea H. Mason, Masuma A. Walji, Elaine J. Lee, and Christine L. MacKenzie. 2001. Reaching movements to augmented and graphic objects in virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) (CHI '01). Association for Computing Machinery, New York, NY, USA, 426–433. <https://doi.org/10.1145/365024.365308>
- [39] Mike McGee, Brian Amento, Patrick Brooks, and Hope Harley. 1997. Fitts and VR: Evaluating Display and Input Devices with Fitts' Law. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 41, 2 (1997), 1259–1262. <https://doi.org/10.1177/1071181397041002119> arXiv:<https://doi.org/10.1177/1071181397041002119>
- [40] Erin A. McManus, Bobby Bodenheimer, Stephan Streuber, Stephan de la Rosa, Heinrich H. Bühlhoff, and Betty J. Mohler. 2011. The Influence of Avatar (Self and Character) Animations on Distance Estimation, Object Interaction and Locomotion in Immersive Virtual Environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (APGV '11). Association for Computing Machinery, New York, NY, USA, 37–44. <https://doi.org/10.1145/2077451.2077458>
- [41] Meta. 2025. Meta Quest MR, VR Headsets & Accessories. <https://www.meta.com/ca/quest>
- [42] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.
- [43] Brad A. Myers, Rishi Bhatnagar, Jeffrey Nichols, Choon Hong Peck, Dave Kong, Robert Miller, and A. Chris Long. 2002. Interacting at a Distance: Measuring the Performance of Laser Pointers and Other Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '02). Association for Computing Machinery, New York, NY, USA, 33–40. <https://doi.org/10.1145/503376.503383>
- [44] Riku Otono, Adélaïde Genay, Monica Perusquia-Hernández, Naoya Isoyama, Hideaki Uchiyama, Martin Hachet, Anatole Lécuyer, and Kiyoshi Kiyokawa. 2023. I'm Transforming! Effects of Visual Transitions to Change of Avatar on the Sense of Embodiment in AR. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 83–93. <https://doi.org/10.1109/VR55154.2023.00024>
- [45] Riku Otono, Naoya Isoyama, Hideaki Uchiyama, and Kiyoshi Kiyokawa. 2022. Third-Person Perspective Avatar Embodiment in Augmented Reality: Examining the Proteus Effect on Physical Performance. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 730–731. <https://doi.org/10.1109/VRW55335.2022.00216>
- [46] Ye Pan and Anthony Steed. 2019. Avatar Type Affects Performance of Cognitive Tasks in Virtual Reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) (VRST '19). Association for Computing Machinery, New York, NY, USA, Article 6, 4 pages. <https://doi.org/10.1145/3359996.3364270>
- [47] Fabian Pointecker, Judith Friedl, Daniel Schwajda, Hans-Christian Jetter, and Christoph Anthes. 2022. Bridging the Gap Across Realities: Visual Transitions Between Virtual and Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 827–836. <https://doi.org/10.1109/ISMAR55827.2022.00101>
- [48] Fabian Pointecker, Judith Friedl-Knirsch, Hans-Christian Jetter, and Christoph Anthes. 2024. From Real to Virtual: Exploring Replica-Enhanced Environment Transitions along the Reality-Virtuality Continuum. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (CHI '24). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3613904.3642844>
- [49] Jose Luis Ponton, Reza Keshavarz, Alejandro Beacco, and Nuria Pelechano. 2024. 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). Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3613904.3642268>
- [50] Sofia Seinfeld, Tiare Feuchtnner, Antonella Maselli, and Jörg Müller. 2021. User Representations in Human-Computer Interaction. *Human-Computer Interaction* 36, 5-6 (Oct. 2021), 400–438. <https://doi.org/10.1080/07370024.2020.1724790>
- [51] Richard Skarbez, Missie Smith, and Mary C. Whitton. 2021. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality* 2 (2021). <https://doi.org/10.3389/frvir.2021.647997>
- [52] Mel Slater and Martin Usoh. 1994. Body centred interaction in immersive virtual environments. *Artificial life and virtual reality* 1, 1994 (1994), 125–148.
- [53] Anthony Steed, Ye Pan, Fiona Zisch, and William Steptoe. 2016. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE virtual reality (VR)*. IEEE, 67–76.
- [54] J Edward Swan, Adam Jones, Eric Kolstad, Mark A Livingston, and Harvey S Smallman. 2007. Egocentric depth judgments in optical, see-through augmented reality. *IEEE transactions on visualization and computer graphics* 13, 3 (2007), 429–442.
- [55] William B. Thompson, Peter Willemsen, Amy A. Gooch, Sarah H. Creem-Regehr, Jack M. Loomis, and Andrew C. Beall. 2004. Does the quality of the computer graphics matter when judging distances in visually immersive environments. *Presence: Teleoper. Virtual Environ.* 13, 5 (Oct. 2004), 560–571. <https://doi.org/10.1162/1054746042545292>
- [56] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Balagopal Raveendranath, Christopher C. Pagano, Andrew C. Robb, Wen-Chieh Lin, and Sabarish V. Babu. 2023. Give Me a Hand: Improving the Effectiveness of Near-field Augmented Reality Interactions By Avatarizing Users' End Effectors. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2412–2422. <https://doi.org/10.1109/TVCG.2023.3247105>
- [57] 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. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2258–2268. <https://doi.org/10.1109/TVCG.2023.3247041> Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- [58] Colin Ware. 1995. Dynamic stereo displays. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 310–316.
- [59] Colin Ware. 2000. *Information visualization: perception for design*. Morgan Kaufmann.
- [60] Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt, Wolfgang Broll, and Alexander Raake. 2023. A systematic review on the visualization of avatars and agents in ar & vr displayed using head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2596–2606.
- [61] Franziska Westermeyer, Larissa Brübach, Carolin Wienrich, and Marc Erich Latoschik. 2024. Assessing Depth Perception in VR and Video See-Through AR: A Comparison on Distance Judgment, Performance, and Preference. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (May 2024), 2140–2150. <https://doi.org/10.1109/TVCG.2024.3372061>
- [62] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (June 1998), 225–240. <https://doi.org/10.1162/105474698565686>
- [63] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 143–146. <https://doi.org/10.1145/1978942.1978963>