Investigating the Effects of Avatarization and Interaction Techniques on Near-field Mixed Reality Interactions with Physical Components

Roshan Venkatakrishnan, Rohith Venkatakrishnan, Ryan Canales, Balagopal Raveendranath, Christopher C. Pagano, Andrew C. Robb, Wen-Chieh Lin, and Sabarish V. Babu

Fig. 1: Third person perspective of participant performing the peg transfer task with the holographic ring being grasped.

Abstract— Mixed reality (MR) interactions feature users interacting with a combination of virtual and physical components. Inspired by research investigating aspects associated with near-field interactions in augmented and virtual reality (AR & VR), we investigated how avatarization, the physicality of the interacting components, and the interaction technique used to manipulate a virtual object affected performance and perceptions of user experience in a mixed reality fundamentals of laparoscopic peg-transfer task wherein users had to transfer a virtual ring from one peg to another for a number of trials. We employed a 3 (Physicality of pegs) X 3 (Augmented Avatar Representation) X 2 (Interaction Technique) multi-factorial design, manipulating the physicality of the pegs as a between-subjects factor, the type of augmented self-avatar representation, and the type of interaction technique used for object-manipulation as within-subjects factors. Results indicated that users were significantly more accurate when the pegs were virtual rather than physical because of the increased salience of the task-relevant visual information. From an avatar perspective, providing users with a reach envelope-extending representation, though useful, was found to worsen performance, while co-located avatarization significantly improved performance. Choosing an interaction technique to manipulate objects depends on whether accuracy or efficiency is a priority. Finally, the relationship between the avatar representation and interaction technique dictates just how usable mixed reality interactions are deemed to be.

Index Terms—Mixed Reality, Self-Avatars, Interactions in MR, Tangible entities

1 Introduction

Mixed Reality (MR) is rapidly gaining popularity and a number of technological conglomerates are actively invested in realizing the technology's ubiquity. Augmented reality (AR) enables the registration of computer-generated interactive virtual content, often in two or three dimensions, onto the physical world [2]. Mixed reality on the other hand furthers this concept by bundling in interactive physical components to the experience. In addition to being able to see overlaid virtual artifacts like in AR, MR allows for physical components to be able to interact with virtual components, essentially breaking down the barriers between physical and virtual realities [48]. However, there are no universally agreed upon standards differentiating AR and MR, thus causing academics, researchers, and technologists to use the terms interchangeably [70].

Interaction constitutes a fundamental aspect of MR applications, where users engage with digitally augmented virtual components that are registered and integrated into the physical environment. These components encompass a wide array of virtual entities, including objects, menus, icons, interfaces, and even virtual humans [2]. While these interactions can occur in both near and far regions, the majority of interactions with virtual entities predominantly manifest in the near field. These interactions can be facilitated through various means, such as the utilization of specialized hardware like controllers or the adoption of intuitive and natural methods like speech, eye tracking, and hand gestures [78]. In contemporary devices, simple selection interactions are often facilitated using a combination of methods like eye gaze and hand gestures wherein users make gestures while looking at the region of interest for selection. While such methods are promising for simple selection tasks, users commonly exhibit a preference for direct hand-based interaction to perform object manipulations, due to their inherent familiarity [34]. Consequently, natural freehand gestures persist as the prevailing standard for manipulating virtual content in close

- *Roshan Venkatakrishnan and Rohith Venkatakrishnan are Research Associates at the University of Florida. E-mails: rvenkatakrishnan@ufl.edu, rohith.venkatakr@ufl.edu*
- *Ryan Canales and Balagopal Raveendranth are PhD Candidates at Clemson University. E-mails: rcanale@clemson.edu, braveen@g.clemson.edu*
- *Christopher Pagano is a Professor in the Department of Psychology at Clemson University. E-mail: cpagano@clemson.edu*
- *Wen-Chieh Lin is a Professor in the Department of Computer Science at National Yang Ming Chiao Tung University. E-mail: wclin@cs.nctu.edu.tw*
- *Andrew C. Robb and Sabarish V. Babu are Associate Professors in the School of Computing at Clemson University. E-mails: arobb@clemson.edu, sbabu@clemson.edu*

Manuscript received 4 October 2023; revised 17 January 2024; accepted 24 January 2024. Date of publication 4 March 2024; date of current version 15 April 2024. This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2024.3372050, provided by the authors. Digital Object Identifier no. 10.1109/TVCG.2024.3372050

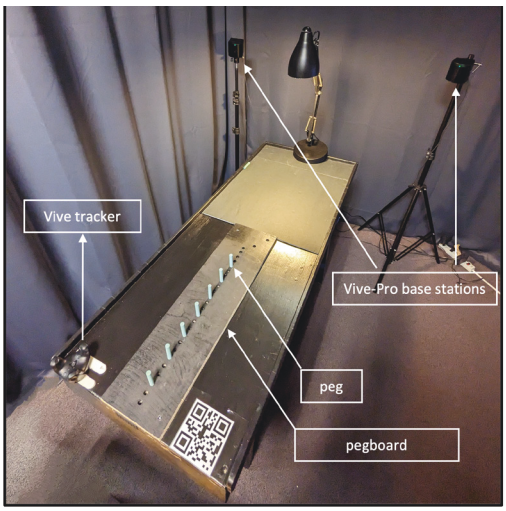


Fig. 2: Physical apparatus: wooden pegboard screwed onto the table with 7 pegs slotted into the holes of the board.

proximity. MR Applications in areas like surgery, electrical circuitry, mechanical assembly, industrial training, etc., require fine motor control and typically feature passive haptics because they involve real world physical entities that provide realistic haptic feedback. These kinds of MR applications are increasingly being utilized, making research on near field interactions occurring within this medium timely and important. While researchers have largely investigated aspects related to perception-action coordination in motor control tasks occurring in virtual reality (VR) with active haptics [6, 8], it remains to be seen as to how such interactions take place in MR applications featuring passive haptics such as those aforementioned.

Researchers have recently shown that near-field, fine-motor AR interactions, requiring precise perception-action coordination are more accurately performed when users are avatarized [74]. The benefit of this avatarization was explained to stem from the visualization of the task relavant information - a visual representation of where the system tracks the user's hand (interacting layer). While avatarizational benefits were observed in the context of AR, it remains to be seen if the same holds true for interactions occurring in MR. Along these lines, researchers are yet to explore the effects of avatarization on users' ability to perform fine-motor, precision-requiring interactions when the interacting entities are comprised of both virtual and physical integrants. Researchers have explored the potential extension of users' reach envelopes by applying translational gains to their end-effectors, as in [8, 15, 82]. Translational gains allow users to interact with objects farther away from them while their physical end-effectors continue to remain in close proximity, also helping to combat tracking issues that may occur when the user's hand moves beyond the tracked region. Translational losses are different from translational gains in that the former moves the virtual end-effector to a smaller extent than the actual end-effector, while the latter does the opposite. In near-field interactions, transitional losses offer more precision than gain-based representations because gains are more sensitive to small movements of the hand. Despite this compromise in performance, gain-based representations are popular due to their utility, making it appropriate for researchers to study if this kind of affordance compromises performance on MR interaction tasks. Furthermore, in contrast to interaction scenarios in AR and VR, it remains to be seen as to how different interaction techniques play out in the context of MR experiences with physical components. In an attempt to contribute to this problem space as a whole, we detail the results of an experimental investigation that appraises how fine-motor MR interactions are affected by different types of end-effector avatarizations, the physicality of interacting components, and the type of interacting technique used for object manipulation.

2 Related Works

2.1 Interaction in Augmented and Mixed Reality

Interaction within the realm of Augmented and Mixed Reality (AR/MR) typically entails interaction with virtual entities such as objects, menus, icons, and the like, which are spatially registered in three dimensions and superimposed onto the real-world environment [2]. These interactions may be facilitated through natural means, including speech, eye gaze, hand gestures, and facial expressions, or through hardware-driven techniques, utilizing hand-held controllers [78]. Researchers have been actively exploring these modalities to foster intuitive and immersive interactions with virtual entities in close proximity [1, 84]. The prevailing consensus from much of this research suggests that users exhibit greater efficiency in performing near-field interactions via direct manipulation, often favoring modalities supported by natural means, particularly freehand gestures [24, 34, 57]. However, it is worth noting that certain studies have reported preferences for non-natural interaction over their natural counterparts [59].

Direct hand-based interactions in AR frequently necessitate users to perform gestures such as tapping markers [12, 49], pinching or pushing with fingertip precision [49, 56], opening their palms or making a fist [58, 61], or maneuvering virtual objects within the spatial environment [14, 64]. A recent investigation comparing three modes of object manipulation interaction in MR shows that while users prefer direct hand gesture-based interactions, their performance is significantly better when using a 'worlds-in-minature' or gaze+pinch based approach [34]. While some systems are restricted to uni-manual (single-handed) interactions [46, 61], others support bi-manual (dual-handed) interactions. Typically, hand-gesture recognition in AR/MR systems is realized either through wearable data gloves [43, 53] or by employing depth cameras, video cameras, or infrared sensors [37, 73]. Despite the former offering superior accuracy, reliability, and haptic feedback potentially, gesture recognition through vision-based systems are often preferred due to simplicity, unrestricted freedom of movement, and the absence of specialized hardware requirements.

Natural hand gestures seem particularly well-suited for near-field interactions, while manipulation of virtual objects situated at considerable distances from the user present unique challenges. For instance, pointing to occluded objects may necessitate nonlinear spatial and visual mapping in noisy environments [17]. Furthermore, far-field interactions may entail manipulation of geometries extending beyond one's arm reach [28, 54]. Research on this front suggests that precision in terms of pointing and selection performance tends to degrade with target size and distance, thus rendering far-field AR interactions supported by natural hand gestures challenging [36]. Consequently, multi-modal interaction techniques amalgamating technologies such as speech, gaze, and gesture recognition, have been introduced and investigated [30, 76]. Research suggests that gaze based selection interactions can be faster than hand-gesture or device based selections in [38]. Similar findings have been obtained on large screen stereoscopic displays wherein gaze based input tends to be faster than hand-pointing [67]. Furthermore, some research indicates that speech outperforms gestures in terms of accuracy, but the simplicity and speed of gestures compensate for potential loss in precision [11]. Essentially, it appears that the challenges associated with freehand gestures predominantly manifest in far-field interactions, where increased distances result in a breakdown of direct manipulation metaphors [33].

Two common gesture-based techniques to grasp and manipulate virtual objects in MR include metaphoric and isomorphic hand interactions [44]. The former relies on image schemas and conceptual metaphors which predicate system responses while the latter revolves around interactions that establish a one-to-one correspondence between users' input actions and the resulting responses. A pinch gesture to manipulate an object through space exemplifies the metaphoric paradigm while a gesture that relies on grasping the virtual object by its boundaries represents an isomorphic approach. Recent research suggests that the isomorphic approach is perceived as more usable and natural while the metaphoric approach may be conducive for resizing tasks [20]. Conversely, other studies have found no significant differences between

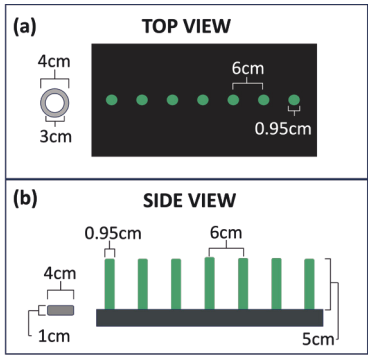


Fig. 3: Schematic representation of the physical apparatus.

these techniques, both in terms of task performance and subjective user experience [66].

In scenarios where precise manipulation of virtual objects is of importance, the chosen input modality profoundly influences user performance. Researchers have thus examined scenarios in which specific input modalities exhibit superior performance in AR tasks. For instance, touch and freehand gestures have proven highly effective for selection tasks involving individual virtual entities, while voice commands excel in tasks related to the creation of new visualizations [3]. Regarding 3D cursor placement tasks, studies have indicated that users often favor handheld controllers, with comparable performance levels when users employ remote controllers and embodied head-tracked cursors [79]. In a recent investigation comparing the impact of various input modalities on Fitts' law-based target selection tasks in AR, the opacity level of the target was found to have negligible effects on performance. However, the study did highlight that a ray-casting based selection technique outperformed both touchpad and gesture-based approaches in terms of throughput and error rates [47]. Despite these findings, near-field interactions in AR and MR that necessitate the selection and manipulation of virtual objects often continue to leverage hand gestures as the de facto standard of interaction given users' preference and familiarity with the same, thus allowing them to learn how to gesturally control virtual entities in relatively short times [60].

2.2 Effects of Avatarization on AR and MR Interactions

The idea of provisioning users with avatars in AR/MR is gaining traction given the potential associated with perceiving embodiment. This paves the way for several applications in the field of education [32], medical training [39], remote collaboration [51], gaming [63], etc. Avatarization in these mediums largely depends on factors like the type of display (handheld or wearable), the user perspective (1PP or 3PP), and the rendering technique employed (optical or video). OST HMDs augment avatars by overlaying content on the real world while VST HMDs use in-painting techniques to modify live camera feed to augment avatars. Another approach involves the use of holographic augmented mirrors to project avatars that users can embody. Using this technique has been shown to influence perceptions of body weight [80]. The extent of avatarization in AR can vary, ranging from scenarios with no virtual elements (where users see their actual bodies) to full avatarization, where users entirely embody and control virtual bodies [21]. Partial avatarization lies in between wherein human limbs are overlaid or replaced with virtual counterparts, thus finding great relevance in medical applications like rehabilitation and prosthetics [26, 69].

Limited work has looked into how these augmented avatars affect interactions. Embodying more muscular self avatars can help improve physical performance [52]. Recently, Venkatakrishnan et al. studied the effects of avatarization on a near-field obstacle avoidance AR interaction task [74]. They compared 2 different end-effector representations against a baseline control group with no avatar in terms of performance and user experience. Both the iconic and human-like avatars were found to improve performance due to the affordance of the task-relevant information (a visualization of where the system tracks the users hand). Avatarization may also have impacts on interactions that extend beyond a user's natural reach envelope. Research on this front has shown that augmenting users with expandable arms is conducive for far field interactions with objects connected to the system [18]. More recently, it was shown that a translational gain applied to a user's end-effector, can extend their workspace but compromise interaction accuracy, and efficiency in VR [8]. In this work, the translational gain was realized such that as the user moved their physical end-effector, their virtual end-effector moved a larger distance (based on a proportion of distance from a predefined origin), allowing the user to reach virtual artifacts located further away from their natural reach envelope. Users in this study were tasked with transferring a virtual ring from one peg to another based on the fundamentals of laproscopic surgery training task [19, 71]. This scenario is apt for the study of mixed reality interactions as it presents researchers with an opportunity to manipulate the physicality/virtuality of the apparatus in conjunction with variations applied on users' self-avatar representations.

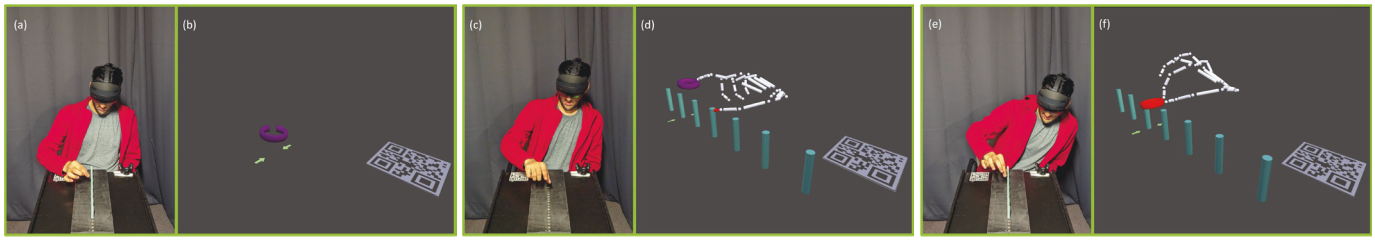
2.3 Impact of Physical Components on MR Interactions

MR experiences are characterized by interactions between virtual and physical entities. Tangible User Interfaces (TUIs) of matching sizes and forms allows to establish a mapping between the virtual and physical content [27]. In terms of MR interactions with physical components, it has been shown that object manipulation interactions are significantly improved when using tangible components [5]. In contrast to VR, aspects of tangibility derived from interacting with physical entities generally allows for improved manipulation [22, 65]. Researchers have also explored how simple objects like cups, cubes, and paddles are manipulated wherein these objects served as controllers for their virtual counterparts [25]. A cube shaped, marker-based tangible interface has been utilized as a proxy for users to rotate a digital brain and make selections [23]. Tangible stylus tools have also been used to slice volumetrically represented data [31]. While such efforts have been examined in the context of interactions involving physical entities, limited work looks at these interactions in near-field tasks that require fine-motor control and precise perception-action coordination. Moreover, avatars are seldom studied in contexts that involve MR interactions with physical entities. We attempt to bridge this gap, contributing the knowledge base of how avatarization and the physicality of interacting entities affect performance on a mixed reality perception-action coordination task, requiring fine motor control and dexterity.

3 System Description

3.1 Physical Apparatus

We constructed a rectangular physical pegboard base using birch-sanded plywood. The board was carefully cut (15.77 cm wide, 61 cm long, and 1.34 cm thick) with a vertical panel saw and then uniformly painted matte black. Using a 9-inch (22.86 cm) bench electric drill press, we drilled holes (radius = 95 mm) in a straight line, centered about the width and evenly spaced along the length of the pegboard (see Figure 2). The distance between the centers of adjacent holes was exactly 2 cm apart. Seven identical cylindrical pegs were crafted by precisely cutting wooden dowels (radius = 5 mm, height = 6.34 cm) using a specialized electrical variable speed scroll saw [77]. We carefully painted each peg a coastal sage matte color, ensuring the coating minimally affected the thickness of the pegs. The pegs slotted into the holes in the wooden board securely enough to prevent movement of the pegs, even when touched physically. The height of the pegs (6.34 cm) was chosen accounting for the thickness of the peg board (1.34 cm), resulting in a final peg height of approximately 5 cm when slotted. We measured the height of each slotted peg 10 times using a vernier caliper to verify the final slotted peg heights for consistency. The mean height of each peg was calculated (49.82 cm, 49.92 cm, 49.96 cm, 49.37 cm, 49.52 cm, 49.39 cm, and 49.30 cm). The wooden peg board was securely mounted on top of an 80 centimeter tall wooden table using screws drilled through both the board and table. The front edge of the peg board was flush with the edge of the table. The pegs were slotted into holes such that the first peg was 7 cm away from the front of the board and the center-to-center distance between successive pegs was 6 cm. A lamp was placed on the far end of the table to facilitate better tracking and viewing of the apparatus when wearing the HMD.



(a) No-Avatar, Physical pegs, Pinch-to-grasp (b) Avatar, Virtual pegs, Stick-on-touch (c) Avatar-Gain, Virtual+Physical pegs, Pinch-to-grasp

Fig. 4: Third person perspective of the experience in the real world (left) and in the MR simulation (right). The images on the right depict the holographic content augmented in the specific conditions. Each sub-figure depicts a particular event during a trial. Sub-figure (2) depicts the ring being centered around a physical destination peg while (4) shows an avatar collision with a virtual peg. Subfigure (6) shows a ring collision with virtual+physical destination peg.

The physical apparatus and its schematic representation are depicted in figures 2 and 3 respectively.

3.2 Hardware and Virtual Components (Holograms)

The simulation was built using Unity 2020.2.2f1 and rendered on a Microsoft Hololens 2 optical-see-through (OST) HMD using a computer equipped with an Intel i7-8700 CPU, 32 GB of RAM, and an NVIDIA RTX 2080 GPU. The Hololens 2 has a 54°diagonal field of view with a frame refresh rate of 60 Hz. The built-in speakers on the device provide a head-related transfer function (HRTF) enabling accurate spatial audio. Users were seated in front of the table with the wooden peg-board base.

The holographic components used for the peg-transfer task (section 4.1) include the virtual ring and virtual pegs. The virtual ring was 0.75 cm thick with an outer diameter of 4 cm and an inner diameter of 3 cm. A uniformly patterned, non-solid white texture was applied to the ring, making its contour salient. The virtual pegs were modeled to match the dimensions of their physical counterparts. The material and color applied to the virtual pegs were matched by sampling an image of the physical pegs taken with a high-quality iPhone 11 camera. For the condition where the pegs were purely physical, a custom shader was applied to the virtual pegs such that they were invisible yet occluded the ring appropriately. This gave the illusion that the physical pegs occluded parts of the virtual ring when seen through the display. In the Virtual+Physical peg condition, the virtual pegs were rendered on top of the physical pegs. Computations were based on the registered positions of the pegs.

3.3 Calibration and Registration Routine

A three-pronged, stepwise calibration routine was used to register the precise location of the virtual pegs on the physical pegboard. First, a simulation was run on a PC equipped with the Vive Pro Tracking system, which was used for initial registration of the virtual apparatus on the physical pegboard. To accomplish this, an HTC Vive tracker was affixed firmly to a specific position on the physical table (see figure 2) and remained in this position throughout the course of the study. Another HTC Vive tracker was attached to the back of the Microsoft Hololens 2 HMD. The positions of these two trackers were obtained and passed to the simulation running in the Microsoft Hololens 2 using a TCP/IP client-server architecture. The position of the virtual apparatus was computed based on the relative positions of the physical pegboard and the Vive tracker, and the position of the user in the virtual space was based on the relative position of the two Vive trackers. Next, the experimenter precisely adjusted the position and orientation of the virtual scene based on feedback from the participant. The magnitude of these adjustments were as little as 1 mm for translation, and 0.1 degrees for rotation. These fine adjustments ensured that the virtual pegs lined up exactly with the holes on the physical pegboard, starting at the third hole from the front. Finally, a visuo-haptic verification step was performed in which the users placed their actual index fingertips on the top of each virtual peg and confirmed that the passive-haptic feedback from the physical peg corresponded precisely with the registered virtual pegs. For conditions where the pegs were completely virtual, matching physical and virtual QR codes served as visual indicators of alignment between the virtual and physical apparatus in place of the virtual/physical pegs. For the condition where the pegs were physical, once registered, the virtual pegs that were visible during the calibration routine were hidden but were still used for collision detection in the experiment. Furthermore, the location of other physical objects in the room remained unchanged to ensure that both the real-world background and the spatial mapping detected by the headset remained constant. In summary, this carefully constructed routine helped ensure that the positions and orientations of the virtual and physical pegs were near-perfect and consistent for all participants across the conditions.

3.4 Hand Avatar Representations

This study investigated two different augmented hand avatar representations the specifics of which are described in this section. An augmented representation involves the provision of a real-time tracked avatar (hologram) based on where the headset tracks the user's actual hand using camera vision. The two augmented avatar representations were compared to a baseline without any augmented avatar, making a total of three hand representations (figure 4).

No Augmented Avatar (No-Av): Users interact without any augmented (visualized) avatar hands. They can see their own real (actual) hands and base their interactions on the same.

Augmented Avatar (Av): Users are provided with an augmented (visualized) avatar hand for interaction. This avatar is represented as a combination of joints and bones, resembling an iconic (skeletal) representation of one's tracked hand. The bones and joints are rendered white to make the avatar salient. The two aforementioned hand representations are identical to those employed in [74].

Augmented Avatar with Translational Gain (AvG): Users interact using an augmented (visualized) hand avatar whose movements are programmed to move twice as much as the user's actual hand along the movement axis (depth axis running from the first peg to the last) in relation to a predefined origin. At the origin, the avatar is co-located with the user's actual hand, while the gain manifests when moved away from the origin along the movement axis. The origin with respect to this translational gain was set to be 10 cm ahead of the first peg, to ensure that the gain is explicitly perceived even when interacting with the first peg. A custom visualizer script was created and used to visualize the avatar hand based on the tracked positions and orientations of the actual hand in relation to the origin and its movement away from it. This amplification of the movement of the visualized end-effector creates a mismatch in the user's actual hand's and the avatar's movement, allowing for users to extend their interaction space beyond their reach envelope. A similar technique has been investigated in a multitude of studies investigating the potential and implications of offsets and gains in near-field interactions [6, 8, 35, 62]. The AvG and the Av's representations are identical, except that the former undertakes a positional offset in relation to the origin at any given time.

Interactions were facilitated based on computations associated with each specific avatar representation. When provisioned with an avatar (Av or AvG), interactions with the ring and the pegs are based on the visualized avatar. Without an avatar however (No-Av), interactions are based on where the HMD tracks the hands to be.

A system evaluation was conducted to compute the positional offset between the augmented avatars and users' real hand position. The distance between the tip of the index finger of the actual hand and the avatar's fingertip was measured across ten samples using a ruler while

resting the actual hand atop a table. This average offset was computed (M=3.7 mm, SD = 0.64). An evaluation of latency and frame rate in all three conditions was conducted using Niehorster et al.'s method [50]. Ten samples of latency and frame rate for simple translational and rotational movements were measured in all conditions. The analysis revealed that the mean frame rate for the different conditions (sampled in the Hololens 2) was measured and found to be stable and approximately equal to 60Hz in each condition. The mean end-to-end latency of the conditions were as follows: No-Av (Pos. lag = 29.16 ms, Ori. lag = 28.33 ms), Augmented-avatar (Pos. lag = 29.58 ms, Ori. lag = 28.75 ms), AvG (Pos. lag = 29.58 ms, Ori. lag = 27.91 ms).

4 Experiment

4.1 Task

A peg-transfer task that serves as a basic training and evaluation tool for hand-eye coordination in laparoscopic surgical training settings [19,71] was used for this study. Users were tasked with manipulating a holographic ring, transferring it from one peg to another, being as accurate and efficient as possible for a number of trials while avoiding collisions with the pegs. A similar task has been used in recent studies investigating the effects of stereoscopic viewing, haptic feedback, and sensory mismatch on near-field fine motor interactions in VR [6–8]. The experiment was divided into six phases each of which had a specific combination of the avatar representation and interaction technique described in 4.2. Each of these phases commenced with three practice trials, allowing users to familiarize themselves with the mechanics associated with grasping/releasing the ring based on the avatar representation and the interaction technique associated with that phase. In the practice trials, the ring was spawned at a predefined location beside the pegs and the first peg always served as the destination for the ring to be placed on. Upon completion of the practice trials, users performed the peg-transfer task for all of the trials in that phase.

At the start of every trial, one peg was identified as the destination peg to place the ring on. Two virtual arrows, one on each side of the peg (left and right), were augmented on the peg board such that they both pointed, as indicators, toward the destination peg in each trial. Users would then grasp and manipulate the ring from the peg it was currently on and place it on the destination peg as accurately as possible. When successfully placed, the ring turned yellow, the two arrows pointing to the peg disappeared, and a trial completion sound was deployed via the HMD. This was further marked by the reappearance of the two arrows on a new destination peg for the next trial. Users then manipulated the ring, moving it to the next indicated destination. This continued until all of the 21 trials in that phase were completed, after which the ring disappeared, effectively marking the completion of said phase.

For all the events that occurred during a trial, multi-modal feedback was provided to allow users to interact with the system effectively and intuitively. Auditory feedback was provided through the HMD's built-in headphones which provide a head-related transfer function (HRTF), thus allowing to simulate accurate spatial sounds from precise locations associated with the scene. Appropriate feedback was provisioned for the grasping/releasing events, collision events, and trial completion events. At rest, the ring was white. When grasped, however, the ring turned purple and a grasping sound was deployed. Similarly, when the ring was released, the ring turned back to white, and a release sound was deployed. Auditory feedback was intentionally added for the grasping events given that users prefer auditory feedback rather than simply having visual feedback when grasping [10]. During a trial, if the ring collided with any of the pegs, the visual feedback involved the ring turning red for as long as the collision was taking place. The auditory feedback associated with these types of collisions involved a ring-collision sound being deployed. Visual and auditory feedback was also provided for collision events involving the pegs and the users' hands/avatars. In phases that provisioned users with an avatar, the specific parts of the avatar (joints and bones) that were colliding with the pegs were highlighted in red for as long as the collision was taking place. This decision to provide fine-grained visual feedback of the avatar collisions provided users with the specificity required to adjust their hand and finger positions based on the feedback provisioned. In phases without an avatar, feedback of the collisions between the tracked hand and the pegs was provided aurally given that there was no avatar to provide visual feedback with. The sounds of collisions with the pegs were modeled as a function of the distance of the user to the peg that was being collided with. The feedback provided was hence tailored to match the specific avatar representation associated with that phase. It was ensured that the sounds associated with the ring collisions with the pegs, hand/avatar collisions with the pegs, the grasping/releasing events, and trial completion were distinct and different from each other. Moreover, the visual and auditory feedback pertaining to any event was presented simultaneously, thus providing users with rich multimodal feedback that was clearly indicative of the different events that transpired during a trial.

4.2 Study Design

To empirically evaluate how the physicality of interactive artifacts affect users' performance in a near-field peg-transfer-based object manipulation task in a mixed reality setting, we employed a 3 (Physicality of pegs) X 3 (Augmented Avatar Representation) X 2 (Interaction Technique) multi-factorial design manipulating the physicality of the pegs as a between-subjects factor across three experimental conditions: (1) Physical or 'P'(no holograms overlaid on physical components); (2) Virtual+Physical or 'V+P' (holograms overlaid on physical components); (3) Virtual or 'V' (only virtual holograms without physical components). Users in each condition performed the peg-transfer task described in section 4.1 for a number of trials over a number of phases in which the avatar representation and the interaction technique (used to grasp and release objects) were manipulated within-subjects.

Two different interaction techniques categorized as Pinch-to-grab and Stick-on-touch were tested in this study. With the former, users could grasp, release, and manipulate the ring by using a simple pinch gesture using their index finger and thumb fingertips. To grasp the ring, the index and thumb fingertips had to come in contact with any portion of the ring after which it could be manipulated. The ring would be released when the user unclasped these two fingers. With the Stick-on-touch technique, users made a pointing gesture and could select the ring by touching it with their index finger's tip. Upon contact, the ring would attach to the tip, allowing it to be manipulated. To release the ring, an opening thumb gesture had to be made and this was designed taking into consideration that the thumb remains the only free finger. The Stick-on-touch technique is inspired by the "Sticky finger" approach described in [4,55] and is easier to manipulate close-by objects, further offering utility in situations with limited space. Applications involving surgical operations, electrical circuitry, and mechanical assembly are examples of such scenarios where collisions of the end-effector with other artifacts may be undesirable, leaving users with less space for object manipulation. Recently, this kind of interaction technique was investigated in a collision-avoidance based object retrieval task where users retrieved targets from an obstacle-filled interaction volume [74], further encouraging its investigation. For each of these interaction techniques, 3 different avatar representations (section 3.4) were tested, making a total of 6 avatar representation-interaction technique combinations (3 avatar representations x 2 interaction techniques). Each of these combinations was blocked into a phase, making a total of 6 phases in the study.

In each phase, users performed the task for a total of 21 trials. With the apparatus comprising 7 pegs, each peg was selected as a destination 3 times, making a total of 21 trials. The order of the destination pegs selected was randomized for all of the 21 trials, ensuring that no two successive trials featured the same destination peg. Each participant performed the peg-transfer task over 6 phases thus accruing up to a total of 126 (3 avatar representations x 2 interaction techniques X 7 pegs X 3 repeats) trials. Within each physicality condition, a balanced Latin square design was adopted to ensure that all possible orders of avatar representations were equally represented and thus counterbalanced. For each participant, it was also ensured that the interaction technique only changed after all the 3 avatar representations for that technique were experienced. This meant that every participant in a given physicality condition experienced one possible order (out of a total of 6

possible orders) of avatar representations twice, once for each level of the interaction technique. Furthermore, the order of the interaction technique was counterbalanced such that half the users experienced the Pinch-to-grasp technique first while the other half experienced the Stick-on-touch technique first. Thus all possible avatar representation-interaction technique combinations were represented equally across all participants assigned to a physicality condition.

4.3 Measures

Error Distance (Accuracy) - For each trial, the distance between the center of the ring and the center of the destination peg was measured. The maximum error distance possible is numerically equal to the inner radius (0.015 m) of the ring, while an error distance of 0 cm corresponds to a perfectly centered ring on the destination peg. A lower error distance corresponds to higher accuracy.

Time on trial (Efficiency) - In each trial, the total time taken from the start of the trial to the end of the trial (when the ring was successfully placed on the destination peg) was computed, and this served as the operational measure of efficiency. The more time on trial, the less efficient users are.

Perceived Usability of interaction - Users' perceived level of usability associated with interactions was measured using the PSSUQ inventory. Counterintuitively, a lower score corresponds to greater perceived usability [40].

4.4 Research Question and Hypotheses

The overarching aim of this study was directed towards answering the following research question: **"How do the aspects of avatarization, the physicality of interacting components, and interaction techniques used affect near-field mixed reality interactions?"** Downstream of this, we were also interested in ascertaining how effectively users can calibrate their performance over time. We operationalize performance based on the measures described in section 4.3 and developed the following hypotheses that reflect the work discussed in section 2:

H1: Interacting with physical pegs will result in lower accuracy.

H2: The avatar-gain representation will perform worse in terms of efficiency and accuracy.

H3: Using the Pinch-to-grasp technique will result in higher accuracy, efficiency, and perceived usability.

H4: Users will perceive interactions to be more usable with an avatar than without.

H5: Users will improve their accuracy and efficiency over trials.

Aspects related to the technical implementation and functioning of the hardware systems strongly determine what effects can be expected. It is expected that provisioning visual information of where the system registers the physical pegs would allow users to be more accurate than without it. With purely physical pegs, users are expected to perform the task simply based on the physical pegs without having on-line visual information associated with the systems' registration of the physical pegs or the 'interacting layer' [74]. Purely physical pegs are hence expected to generally result in lower accuracies. Regarding avatar representations, prior research suggests that translational gains applied to virtual end-effectors negatively affect performance in terms of efficiency and accuracy [6, 8]. Recent work suggests that co-located augmented avatars yield superior performance and are perceived as more usable [74]. In terms of interaction techniques, the pinch technique is more intuitive than the stick technique given users' familiarity with the same. The physicality of the pegs is further expected to moderate the effects of interaction technique on accuracy. When there is occlusion from physical pegs, the interaction technique being used will have a smaller influence on accuracy than when the pegs are virtual. This is because virtual pegs will not suffer from hand-tracking limitations due to occlusion. Finally, calibration or learning occurs in tasks that involve perception-action coordination in XR [13, 75].

4.5 Participants

A total of 36 participants were recruited for this Institutional Review Board (IRB) approved study, with 12 allotted per physicality condition. This fulfilled the balanced Latin square design ensuring equal representation of the interventions across the conditions (see section 4.2). Given that each participant performed 126 peg transfer trials, this led to a total of 4536 trials of peg-transfer-based object manipulations for analysis. Participant ages ranged from 19 to 45 years old (M = 25.47, SD = 4.74); 17,18, and 1 of whom identified as female, male, and non-conforming respectively. All participants were right-handed and had normal/corrected-to-normal vision. Overall, AR/MR experience did not significantly differ across conditions.

4.6 Procedure

Participants were first greeted and asked to read and sign a consent form (informed consent). After consenting, participants filled out a demographics questionnaire.Following this, participants' arm lengths, interpupillary distances (IPD), and stereo acuity were measured. Participants were then randomly assigned to one of the three experimental conditions. The experimenter then detailed the task they would be performing in the study (see section 4.1), explaining the logistics involved with progressing through the six phases of the experiment. Participants then donned the HMD after which an eye calibration routine was run to customize the viewing experience, allowing for optimal hologram interaction. Then the peg calibration routine described in 3.3 was carried out to establish near-perfect co-location of the virtual and the physical apparatus. Once calibrated, participants proceeded to perform the peg-transfer task in each phase, one after the other.

In each phase, the experimenter explained the mechanics of interactions, demonstrating the gestures required to grasp and release the ring with their avatar representation and interaction technique specific to that phase. After each phase, participants filled out the PSSUQ questionnaire and were allowed to take a break before proceeding to the next. Upon completion, participants removed the HMD and were debriefed and compensated for their time. On average, it took an hour to complete the whole procedure.

5 Results

The error distance and time on trial were the dependent variables considered for analyses. Since repeated measures of each dependent variable were considered for each participant, variables had considerable nesting. As the variables were measured over multiple time steps for each participant, a portion of the variance in each dependent variable can be attributed to a common source – the participant themself. Level 1 (within-participant) variables represent those that change between trials. Level 2 (between-participant) variables represent those that change from participant to participant (the condition). To properly account for variance between and within subjects, Hierarchical Linear Modeling was used [29]. Prior to conducting the analysis, the extent of nesting in the data was assessed by computing the intraclass correlation coefficient (ICC) from the null model for each dependent variable separately. The ICC was calculated to be 0.12 for the error distance indicating that approximately 12% of the variance in error distance was associated with the participant and that the assumption of independence was violated. Similarly, the ICC was calculated to be 0.06 for time on trial. Following a multilevel modeling technique is ideal in these cases. For the analysis of each dependent variable, an initial main effects model was run, such that all main effects (Level 1 and Level 2) were included in the analysis at once. Results for each of these main effects are reported from the initial main effects model. To analyze the interaction, the interaction term was added to the main effects model. The effect size for each fixed effect is presented as the change in R^2 (proportion of variance explained) comparing the model that includes the effect and the same model with the effect removed. The resulting sr^2 (semi-partial r^2) is the percentage of variance uniquely accounted for by the fixed effect [68]. For all the models in the analyses, the only random effect computed was the intercept based on the Participant ID. In this section, the block trial number represents the trial number within a phase (block) and hence ranges from 1 to 21 given that each phase comprised 21 trials. The overall trial number, however, represents the trial number regardless of the phase and thus runs from 1 to 126.

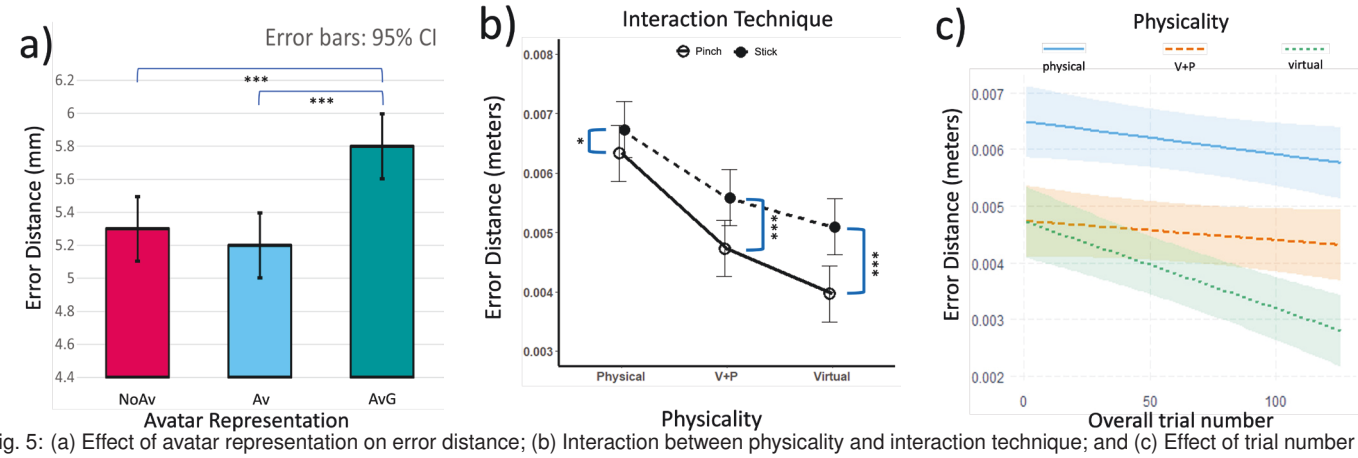


Fig. 5: (a) Effect of avatar representation on error distance; (b) Interaction between physicality and interaction technique; and (c) Effect of trial number on error distance, moderated by physicality. Error bars and shading around each line indicates 95% confidence intervals.

5.1 Accuracy (Error distance)

A linear mixed effects model was run to assess the effects of physicality, avatar representation, interaction technique, overall trial number, and block trial number on error distance. This model with only the main effects (AIC = -39771.28, df = 10) offered a significantly better fit to the data than did the null model (AIC = -39613.52, df = 3), $\Delta\chi^2(7) = 171.76$, $p < 0.001$. The model explained 14% of the variance in error distance (conditional $R^2 = 0.143$, marginal $R^2 = 0.095$). There was a significant effect of physicality on error distance, F(2, 33) = 21.85, $p < 0.001$, $sr^2 = 0.07$. Error distance was significantly larger in the physical condition (M = 0.0065, SE = 0.0002) as compared to the V+P condition (M = 0.0052, SE = 0.0002), t = 4.43, $p < 0.001$ and the virtual condition (M = 0.0045, SE = 0.31), t = 6.46, $p < 0.001$. Error distance was not different between the virtual and V+P condition. There was a significant effect of avatar representation on error distance, F(2, 4495) = 16.11, $p < 0.001$, $sr^2 = 0.006$. Error distance was significantly more for the AvG (M = 0.0058, SE = 0.0001) as compared to the No-Av (M = 0.0053, SE = 0.0001), t = 4.59, $p < 0.001$, as well as the Av representation (M = 0.0052, SE = 0.0001), t = 5.19, $p < 0.001$. There was no significant difference between the Av and the No-Av representation in terms of error distance. There was also a significant main effect of interaction technique, F(1, 4495) = 79.34, $p < 0.001$, $sr^2 = 0.015$. Error distance was greater when using the stick (M = 0.0058, SE = 0.0001) as compared to the pinch (M = 0.0050, SE = 0.0001). The overall trial number also significantly affected the error distance, F(1, 4495) = 8.83, $p = 0.003$, $sr^2 = 0.0008$. As the trial number increased by 1 standard deviation (SD) units, the error distance decreased by 0.3e-5 SD units. Similarly, the block trial number also had a significant effect on the error distance, F(1, 4495) = 24.78, $p < 0.001$, $sr^2 = 0.005$. As the block trial number increased by 1 SD unit, the error distance decreased by 0.4e-4 SD units.

There was a significant interaction between interaction technique and physicality, F(2, 4493) = 5.74, $p = 0.003$, $sr^2 = 0.002$ (figure 5b). When testing simple effects, when participants were in the physical condition, the error distance was significantly more when the stick interaction technique was used (M = 0.0067, SE = 0.0002) as compared to when the pinch technique was used (M = 0.0063, SE = 0.0002), t(1512) = 2.28, $p = 0.02$. When participants were in the V+P condition, the error distance was significantly more when the stick interaction technique was used (M = 0.0055, SE = 0.0002) as compared to when the pinch technique was used (M = 0.0047, SE = 0.0002), t(1512) = 5.72, $p < 0.001$. Similarly, when participants were in the virtual condition, the error distance was significantly more when the stick interaction technique was used (M = 0.0051, SE = 0.0002) as compared to when the pinch technique was used (M = 0.0039, SE = 0.0002), t(1512) = 8.06, $p < 0.001$.

Physicality significantly moderated the effect of trial number on error distance, F(2, 4493) = 8.94, $p < 0.001$, $sr^2 = 0.003$. That is, physicality altered the slope (or rate of change) of the relationship between trial number and error distance. As seen in figure 5c, a test of simple slopes revealed that the simple slope for trial number was negative and significantly different from zero only for the V condition (B = -0.98e-5, SE = 0.2e-5, t = -4.60, $p < 0.001$), while they were not significantly different from zero for the V+P and P conditions.

5.2 Efficiency (Time on trial)

A linear mixed effects model was run to assess the effects of physicality, avatar representation, interaction technique, overall trial number, and block trial number on the time on trial. This model with only the main effects (AIC = 29939.63, df = 10) offered a significantly better fit to the data than did the null model (AIC = 30374.69, df = 3), $\Delta\chi^2(7) = 449.06$, $p < 0.001$. The model explained 14.6% of the variance in time on trial (conditional $R^2 = 0.146$, marginal $R^2 = 0.095$). There was no significant effect of physicality on time on trial. However, there was a significant effect of avatar representation on time on trial, F(2, 4495) = 14.79, $p < 0.001$, $sr^2 = 0.006$. Time on trial was significantly more for the AvG (M = 9.12, SE = 0.31) as compared to the No-Av representation (M = 8.12, SE = 0.31), t = 4.24, $p < 0.001$, as well as the Av representation (M = 7.92, SE = 0.31), t = 5.07, $p < 0.001$. There was no significant difference between the Av representation and No-Av representation in terms of time on trial. There was also a significant main effect of interaction technique, F(1, 4495) = 61.68, $p < 0.001$, $sr^2 = 0.012$. Time on trial was more when using the pinch technique (M = 9.14, SE = 0.297) as compared to the stick technique (M = 7.63, SE = 0.297). The overall trial number also significantly affected the time on trial, F(1, 4495) = 273.49, $p < 0.001$, $sr^2 = 0.04$. As the trial number increased by 1 standard deviation (SD) units, the time on trial decreased by 0.04 SD units. Similarly, the block trial number also had a significant effect on time on trial, F(1, 4495) = 102.15, $p < 0.001$, $sr^2 = 0.02$. As the block trial number increased by 1 SD unit, the time on trial decreased by 0.16 SD units.

There was a significant interaction between interaction technique and avatar representation, F(2, 4493) = 7.32, $p < 0.001$, $sr^2 = 0.003$ (figure 6a). When testing simple effects, when participants were provisioned with the Av representation, time on trial was significantly more when the pinch interaction technique was used (M = 8.81, SE = 0.35) as compared to when the stick technique was used (M = 7.03, SE = 0.35), t(1512) = 5.65, $p < 0.001$. When participants were provisioned with the AvG representation, time on trial was significantly more when the pinch interaction technique was used (M = 10.25, SE = 0.35) as compared to when the stick technique was used (M = 7.99, SE = 0.35), t(1512) = 6.37, $p < 0.001$. However, when participants were provisioned with the No-Av representation, time on trial was not significantly different in the stick interaction technique as compared to the pinch technique.

Physicality was a significant moderator for the effect of trial number on time on trial, F(1, 4493) = 13.36, $p < 0.001$, $sr^2 = 0.005$. That is, physicality altered the slope (or rate of change) of the relationship between trial number and time on trial. As seen in figure 6b, a test of simple slopes revealed that for each physicality, the simple slope for trial number was negative and significantly different from zero. The

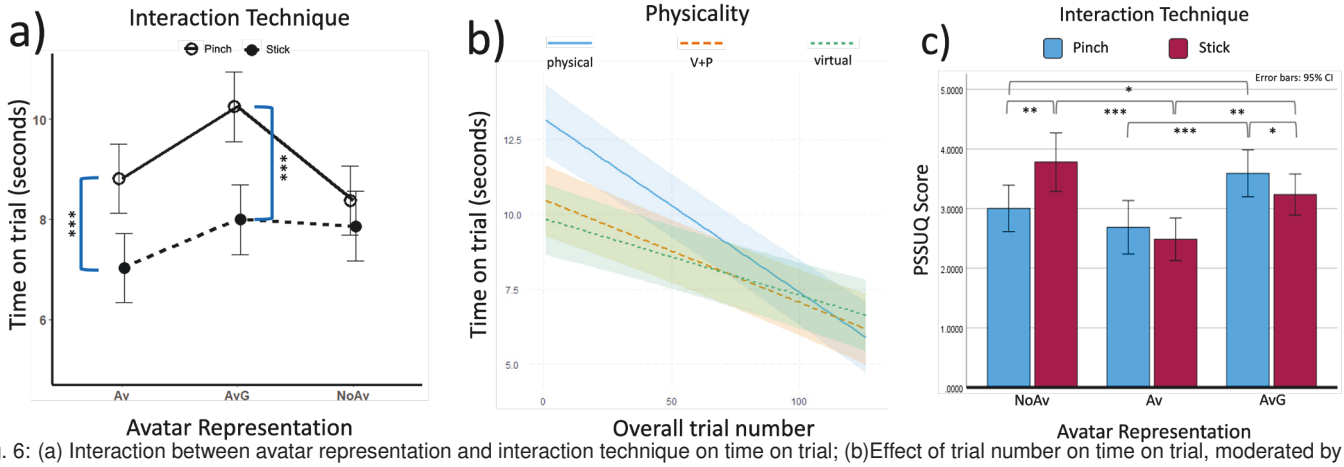


Fig. 6: (a) Interaction between avatar representation and interaction technique on time on trial; (b)Effect of trial number on time on trial, moderated by physicality. Shading around each line indicates 95% confidence intervals. (c) Interaction effect of avatar representation and interaction technique on perceived usability. A lower PSSUQ score corresponds to a higher perceived usability associated with interactions.

physical condition (B = -0.058, SE = 0.005, t = -12.61, $p < 0.001$) had a steeper positive slope as compared to the V+P condition (B = -0.034, SE = 0.005, t = -7.46, $p < 0.001$), and the virtual condition (B = -0.026, SE = 0.005, t = -5.58, $p < 0.001$).

5.3 Perceived Usability of interaction

A Two-way repeated measures ANOVA was conducted to determine the effects of the avatar representation and interaction technique on users' perceived usability. A significant main effect of avatar representation was found, F(2, 68) = 16.105, $p < 0.001$, $\eta_p^2 = 0.321$. The mean PSSUQ score was significantly lower for the Av (M = 2.58, SE = 0.169) as compared to the No-Av (M = 3.39, SE = 0.172), $p < 0.001$, and AvG representation (M = 3.412, SE = 0.163), $p < 0.001$. No other significant differences were found between the representations. There was no significant main effect of the interaction technique F(1, 34) = 0.299, $p = 0.588$, $\eta_p^2 = 0.009$. However, a significant interaction effect between avatar representation and interaction technique was found, F(2, 68) = 8.579, $p < .001$, $\eta_p^2 = 0.201$. As seen in figure 6(c), without an avatar (No-Av), the mean PSSUQ score was significantly higher for the stick (M = 3.77, SE = 0.242) as compared to the pinch technique (M = 3.3002, SE = 0.192), $p < 0.01$. For the AvG representation, the mean PSSUQ score was significantly higher when using the pinch (M = 3.589, SE = 0.194) as compared to the stick technique (M = 3.234, SE = 0.170) $p < 0.05$. No other significant differences were found.

6 Discussion

The statistical analyses pertaining to users' accuracy revealed that purely physical pegs were associated with the highest error distance in comparison to pegs that had some degree of virtuality (V or V+P), thus supporting hypothesis **H1**. This suggests that when the pegs are purely physical, users are more inaccurate in terms of how well they are able to center the virtual ring on pegs. Virtual pegs make the visual information of relevance to the task (registered position of pegs) more salient, thereby allowing for increased accuracy through better online control and perception-action coordination [16, 83]. This idea aligns directly with research suggesting that visualizing the task-dependent information (the interacting layer corresponding to the registered pegs in this case) improves near-field interaction performance [74]. We also found a main effect of the avatar representation, such that the avatar with a translational gain was most inaccurate. It is possible that this result was observed because the gain function while potentializing the extension of one's reach and workspace, creates a mismatch between the visual representation of the avatar and the proprioceptive information pertaining to one's actual hand. This idea is supported by prior work suggesting that adding such gains to the virtual end-effector affects task performance negatively in similar scenarios [8].

In terms of interaction techniques, we found a main effect showing that the Stick-on-touch resulted in higher error distances than the Pinch-to-grasp technique. This can be explained based on the fundamentals of human hand morphology and how digits work in conjunction/isolation to perform precise fine motor tasks. Research suggests that precision with respect to grasping and manipulating objects is achieved using the forces of opposition that are provided by the thumb in humans and chimpanzees [9, 41, 42, 45]. Without opposition, the Stick technique requires users to make use of their index finger in isolation. This lack of the opposable thumb to the index finger could help explain why users were less accurate in terms of centering the ring around the pegs. Interestingly, we found an interaction effect between the physicality and interaction technique showing that the magnitude of differences between the techniques reduces as the virtuality of the pegs reduces. In other words, increases in the degree of physicality (moving from purely virtual to purely physical), seem to make the discrepancy between the two interaction techniques less prominent as evinced in figure 5b. This is possibly because physical pegs can occlude portions of the users' hands, potentially making them less accurate in centering the ring regardless of the interaction technique being used. If this was indeed the reason for this observation, both the V+P and the P conditions should not have differed in terms of the magnitude of differences between the two techniques. Given that this was not observed, it may just be that a combined effect of a decreased salience of the task-relevant visual information (registered position of the pegs), and impoverished end-effector tracking resulting from the physicality of the pegs were interacting to produce the observed results.

We also observed that users significantly improved their accuracy over trials. However, this effect of calibrating accuracy was significantly more pronounced for the V condition as compared to the V+P and P conditions as shown in 5c. This implies that users significantly improved their centering of the ring on the destination pegs over trials when the pegs were virtual rather than when there was any physicality associated with the pegs. In the P condition, the visual information of closest relevance to the task's target object are the physical pegs themselves. Once registered and calibrated, users perform the task simply based on the visual information afforded by the physical pegs assuming that the positional calibration of the pegs is perfect and constant, without any tracking errors like drift and jitter. It is reasonable to expect discrepancies between the registered and actual locations of the pegs [72]. Though the exact same calibration routine was employed for all physicality conditions, visualization of the information of relevance (the registered pegs) in the virtual conditions provides users with an online representation of this potentially discrepant information. This is probably why users were unable to calibrate their accuracy over trials when the pegs were purely physical. This increased salience of the task-relevant information in conditions with some degree of virtuality (V or V+P) may have further aided in calibrating performance over trials. Interstingly, with the V+P condition, users were significantly more accurate than the P condition but didn't improve at a rate similar to the V condition. It is likely the case that seeing virtual pegs overlaid on physical pegs could contribute to providing users with small

degrees of competing information (by virtue of having both physical and virtual information), especially given the registration capabilities of contemporary OST HMDs. These are noteworthy findings given the recent growth and interest in facilitating mixed reality interactions. With contemporary interactive MR training simulations increasingly featuring physical components, it deserves noting that accuracy could be compromised as a result of the lack of the added (required) visual information of relevance to the experience.

We found that users took significantly more time to transfer the ring from one peg to the other when provisioned with the Avatar-gain representation. This is understandable given that correctly manipulating the ring required high levels of precision, an aspect the gain condition may not be conducive for, due to reasons explained previously. Research conducted on a similar task, albeit on a stereoscopic display, has found similar results of degraded efficiency when using a gain [8], directly aligning with our findings. The results observed on efficiency and accuracy with respect to the Avatar-Gain representation offered support for hypothesis **H2**. We also observed a main effect of the interaction technique used to manipulate the ring on users' efficiency. In general, users took less time when using the Stick-on-touch technique to manipulate and place (release) the object on the destination pegs than when using the Pinch-to-grasp technique thus rejecting hypothesis **H3** in terms of efficiency. This is interesting given the fact that users were significantly more accurate in performing the task of placing the ring on the pegs. Another factor that may have contributed to this result may revolve around the relative difficulty associated with grasping the ring. With the Stick technique, all a user had to do to grasp the ring was to establish contact with the ring using their index finger's tip. In comparison, the Pinch-to-grasp technique necessitates the establishment of two contact points of the user's end-effector, namely the index and thumb's tips, to a specific point on the ring. Given the added precision and finger dexterity required to grasp the ring using the latter technique, it is not surprising that users took more time to complete trials when pinching.

We observed an interaction effect suggesting that the avatar representation provided to users moderated the effect of the interaction technique on users' efficiency. When provisioned with an avatar (Av or AvG), users took significantly more time to complete the trials when using the Pinch interaction technique. However, without an avatar, there was no difference between the interaction techniques used to manipulate the object as seen in figure 6a. This is an interesting finding that seems to suggest that without an avatar, users seem to be equally fast at performing the task regardless of the interaction technique used. From a qualitative standpoint, nine users mentioned that the avatars sometimes made them behave in ways that they normally wouldn't, especially when having to adjust the position of their hands, fingers, and tips in order to manipulate the ring. It is possible that some degree of the proteus effect [81] may have been at play when users were provisioned with an avatar, but future research is required to further our understanding of the exact reasons for the occurrence of this effect. Finally, we also observed an effect of calibration of efficiency over trials. Users in all conditions of physicality were able to improve their efficiency in the peg-transfer task, partially supporting hypothesis **H5**. The interaction effect suggested that the rate of calibration was highest in the physical condition in comparison to conditions that feature some degree of virtuality of the pegs as shown in figure 6b. This is understandable given that the efficiency at the start of the experiment was lowest for the physical condition. As trials progressed, however, users in this condition were able to improve their efficiency to become equally adept at completing the trials within similar time frames.

Results on the perceived levels of usability associated with interactions shed some light on user perceptions and the experience. In general, users perceived interactions to be more usable with a colocated avatar that best approximated the system's tracking of their actual hands. In contrast, users found the No-Av and the Avatar-Gain representations to be less usable, thus partially supporting hypothesis **H4**. These results validate the results obtained from a previous study [74], suggesting merit in visualizing the interacting layer in the form of a colocated self-avatar. While we did not observe any main effect of interaction technique, we discovered a fascinating interaction effect between the avatar representation and the interaction technique used to manipulate the ring (figure 6c). With an avatar, there does not seem to be much of a difference between using the Pinch-to-grasp or Stick-on-touch technique from a usability standpoint. However, without an avatar, users find object interactions with pinching to be significantly more usable than a sticky finger approach. These results together offer partial support for hypothesis **H3** in terms of usability. It hence comes as no surprise that contemporary MR devices continue to facilitate interactions in the near field using a pinch technique rather than single finger isolation-based techniques, especially because avatars are seldom provided in mixed and augmented reality settings. Though users take less time when using the latter technique, they are less accurate and more than anything, perceive this form of interaction as less usable.

Taken together, it seems appropriate for MR developers and designers to consider the target requirements of an application when determining how to represent users' end-effectors to facilitate conducive interactions with physical components. Along these lines, providing users with a virtual representation of the interacting layer in the form of a co-located self-avatar seems to benefit most aspects like accuracy, efficiency, and usability associated with near-field interactions. In terms of physicality, a decreased salience of the visual information of relevance associated with purely physical components makes interactions challenging in experiences involving both virtual and physical entities. If accuracy is crucial, it behooves developers to visualize the interacting layer corresponding to the registered positions of physical components that form the central part of the MR experience. With respect to interaction techniques, there appear to be different situations that merit either a pinch-based or stick-based approach. The former is suitable when higher levels of accuracy are desired while the latter seems to lend itself more towards improvements in speed while trading off accuracy. These results are particularly useful for MR surgical and industrial training applications where users are required to carefully manipulate objects like surgical tools, electrical circuitry, and other devices. Our findings highlight the need for application designers to factor in the desired qualities of the experience, allowing them to tailor an appropriate end-effector representation and interaction technique based on the application being designed.

7 Conclusion and Future Work

In this work, we investigated how avatarization and the physicality of interacting components affect users' performance in a near-field interaction task in MR. Users were tasked with carefully transferring a holographic ring from one peg to another for a number of trials while being as accurate and efficient as possible. We employed a multi-factorial design manipulating the physicality of the pegs between participants. We further examined the effects of avatar representations and interaction techniques. Results indicated that users were significantly more accurate when the pegs were virtual than when they were physical given a higher salience in the visual information of relevance associated with the task. From an avatar perspective, providing users with gain-based representations negatively affects performance while provisioning them with co-located avatars tend to significantly improve performance. Settling on an interaction technique to manipulate objects would depend on whether accuracy or efficiency is of priority. Finally, the relationship between the avatar representation and interaction technique dictates just how usable mixed reality interactions are deemed to be.

In future work, we wish to investigate if and how avatarization differentially affects performance when manipulating objects of different physicalities. While this study focused solely on manipulating virtual objects, it remains to be seen how the relationship between the target and manipulated object's physicalities affects interactions in MR. Exploring these phenomena in video see-through MR experiences potentiates several avenues for research that future innovators will collectively draw a richness of wealth and knowledge from.

Acknowledgments

The authors would like to thank the participants of our studies for their time and effort. This work was supported in part by the US National Science Foundation (CISE IIS HCC) under Grant No. 2007435.

REFERENCES

[1] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013. 2

[2] R. T. Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997. 1, 2

[3] S. K. Badam, A. Srinivasan, N. Elmqvist, and J. Stasko. Affordances of input modalities for visual data exploration in immersive environments. In *2nd Workshop on Immersive Analytics*, 2017. 3

[4] D. A. Bowman. *Interaction techniques for common tasks in immersive virtual environments: design, evaluation, and application*. Georgia Institute of Technology, 1999. 5

[5] E. Bozgeyikli and L. L. Bozgeyikli. Evaluating object manipulation interaction techniques in mixed reality: Tangible user interfaces and gesture. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 778–787. IEEE, 2021. 3

[6] D. Brickler and S. V. Babu. An evaluation of screen parallax, haptic feedback, and sensory-motor mismatch on near-field perception-action coordination in vr. *ACM Transactions on Applied Perception (TAP)*, 18(4):1–16, 2021. 2, 4, 5, 6

[7] D. Brickler, R. J. Teather, A. T. Duchowski, and S. V. Babu. A fitts' law evaluation of visuo-haptic fidelity and sensory mismatch on user performance in a near-field disc transfer task in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 17(4):1–20, 2020. 5

[8] D. Brickler, M. Volonte, J. W. Bertrand, A. T. Duchowski, and S. V. Babu. Effects of stereoscopic viewing and haptic feedback, sensory-motor congruence and calibration on near-field fine motor perception-action coordination in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 28–37. IEEE, 2019. 2, 3, 4, 5, 6, 8, 9

[9] I. M. Bullock, T. Feix, and A. M. Dollar. Human precision manipulation workspace: Effects of object size and number of fingers used. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5768–5772. IEEE, 2015. 8

[10] R. Canales and S. Jörg. Performance is not everything: Audio feedback preferred over visual feedback for grasping task in virtual reality. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pp. 1–6, 2020. 5

[11] Z. Chen, J. Li, Y. Hua, R. Shen, and A. Basu. Multimodal interaction in augmented reality. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 206–209. IEEE, 2017. 2

[12] K.-Y. Cheng, R.-H. Liang, B.-Y. Chen, R.-H. Laing, and S.-Y. Kuo. icon: utilizing everyday objects as additional, auxiliary and instant tabletop controllers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1155–1164, 2010. 2

[13] E. Ebrahimi, A. Robb, L. S. Hartman, C. C. Pagano, and S. V. Babu. Effects of anthropomorphic fidelity of self-avatars on reach boundary estimation in immersive virtual environments. In *Proceedings of the 15th ACM Symposium on Applied Perception*, pp. 1–8, 2018. 6

[14] J. Ehnes. A tangible interface for the ami content linking device—the automated meeting assistant. In *2009 2nd Conference on Human System Interactions*, pp. 306–313. IEEE, 2009. 2

[15] S. Esmaeili, B. Benda, and E. D. Ragan. Detection of scaled hand interactions in virtual reality: The effects of motion direction and task complexity. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 453–462. IEEE, 2020. 2

[16] B. R. Fajen. *Visual control of locomotion*. Cambridge University Press, 2021. 8

[17] A. O. S. Feiner. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proc. UIST*, vol. 3, pp. 81–82, 2003. 2

[18] T. Feuchtner and J. Müller. Extending the body for interaction with reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5145–5157, 2017. 3

[19] G. M. Fried. Fls assessment of competency using simulated laparoscopic tasks. *Journal of Gastrointestinal Surgery*, 12:210–212, 2008. 3, 5

[20] M. Frutos-Pascual, C. Creed, and I. Williams. Head mounted display interaction evaluation: manipulating virtual objects in augmented reality. In *IFIP Conference on Human-Computer Interaction*, pp. 287–308. Springer, 2019. 2

[21] A. C. S. Genay, A. Lécuyer, and M. Hachet. Being an avatar" for real": a survey on virtual embodiment in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 2021. 3

[22] L. Gerini, F. Solari, and M. Chessa. A cup of coffee in mixed reality: analysis of movements' smoothness from real to virtual. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 566–569. IEEE, 2022. 3

[23] S. R. Gomez, R. Jianu, and D. H. Laidlaw. A fiducial-based tangible user interface for white matter tractography. In *Advances in Visual Computing: 6th International Symposium, ISVC 2010, Las Vegas, NV, USA, November 29–December 1, 2010, Proceedings, Part II 6*, pp. 373–381. Springer, 2010. 3

[24] T. Ha, S. Feiner, and W. Woo. Wearhand: Head-worn, rgb-d camera-based, bare-hand user interface with visually enhanced depth perception. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 219–228. IEEE, 2014. 2

[25] T. Ha and W. Woo. An empirical evaluation of virtual hand techniques for 3d object manipulation in a tangible augmented reality environment. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 91–98. IEEE, 2010. 3

[26] C. Heinrich, M. Cook, T. Langlotz, and H. Regenbrecht. My hands? importance of personalised virtual hands in a neurorehabilitation scenario. *Virtual Reality*, 25(2):313–330, 2021. 3

[27] A. Hettiarachchi and D. Wigdor. Annexing reality: Enabling opportunistic use of everyday objects as tangible proxies in augmented reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1957–1967, 2016. 3

[28] T. N. Hoang, R. T. Smith, and B. H. Thomas. Ultrasonic glove input device for distance-based interactions. In *2013 23rd International Conference on Artificial Reality and Telexistence (ICAT)*, pp. 46–53. IEEE, 2013. 2

[29] D. A. Hofmann. An overview of the logic and rationale of hierarchical linear models. *Journal of management*, 23(6):723–744, 1997. 6

[30] H. Ishiyama and S. Kurabayashi. Monochrome glove: A robust real-time hand gesture recognition method by using a fabric glove with design of structured markers. In *2016 IEEE virtual reality (VR)*, pp. 187–188. IEEE, 2016. 2

[31] P. Issartel, F. Guéniat, and M. Ammi. Slicing techniques for handheld augmented reality. In *2014 IEEE symposium on 3D user interfaces (3DUI)*, pp. 39–42. IEEE, 2014. 3

[32] A. S. Johnson and Y. Sun. Spatial augmented reality on person: Exploring the most personal medium. In *International Conference on Virtual, Augmented and Mixed Reality*, pp. 169–174. Springer, 2013. 3

[33] E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th international conference on Multimodal interfaces*, pp. 12–19, 2003. 2

[34] H. J. Kang, J.-h. Shin, and K. Ponto. A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 275–284. IEEE, 2020. 1, 2

[35] K. Kohm, J. Porter, and A. Robb. Sensitivity to hand offsets and related behavior in virtual environments over time. *ACM Transactions on Applied Perception*, 19(4):1–15, 2022. 4

[36] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010. 2

[37] P. Kyriakou and S. Hermon. Can i touch this? using natural interaction in a museum augmented reality system. *Digital Applications in Archaeology and Cultural Heritage*, 12:e00088, 2019. 2

[38] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018. 2

[39] E. Lamounier Jr, K. Lopes, A. Cardoso, and A. Soares. Using augmented reality techniques to simulate myoelectric upper limb prostheses. *Journal of Bioengineering & Biomedical Science*, 1:010, 2012. 3

[40] J. R. Lewis. Psychometric evaluation of the pssuq using data from five years of usability studies. *International Journal of Human-Computer Interaction*, 14(3-4):463–488, 2002. 6

[41] K. Li, R. Nataraj, T. L. Marquardt, and Z.-M. Li. Directional coordination of thumb and finger forces during precision pinch. *PloS one*, 8(11):e79400, 2013. 8

[42] Z.-M. Li and J. Tang. Coordination of thumb joints during opposition. *Journal of biomechanics*, 40(3):502–510, 2007. 8

[43] G. Lu, L.-K. Shark, G. Hall, and U. Zeshan. Immersive manipulation of virtual objects through glove-based hand gesture interaction. *Virtual Reality*, 16(3):243–252, 2012. 2

[44] A. Macaranas, A. N. Antle, and B. E. Riecke. What is intuitive interaction?

balancing users' performance and satisfaction with natural user interfaces. *Interacting with Computers*, 27(3):357–370, 2015. 2

[45] M. W. Marzke. Precision grips, hand morphology, and tools. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 102(1):91–110, 1997. 8

[46] D. Merrill and P. Maes. Augmenting looking, pointing and reaching gestures to enhance the searching and browsing of physical objects. In *International Conference on Pervasive Computing*, pp. 1–18. Springer, 2007. 2

[47] D. M. Mifsud, A. S. Williams, F. Ortega, and R. J. Teather. Augmented reality fitts' law input comparison between touchpad, pointing gesture, and raycast. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 590–591. IEEE, 2022. 3

[48] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, vol. 2351, pp. 282–292. Spie, 1995. 1

[49] T. Nagel and F. Heidmann. Exploring faceted geo-spatial data with tangible interaction. *GeoViz 2011*, pp. 10–11, 2011. 2

[50] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017. 5

[51] S. Noh, H.-S. Yeo, and W. Woo. An hmd-based mixed reality system for avatar-mediated remote collaboration with bare-hand interaction. In *ICAT-EGVE*, pp. 61–68, 2015. 3

[52] R. Otono, N. Isoyama, H. Uchiyama, and K. Kiyokawa. 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)*, pp. 730–731. IEEE, 2022. 3

[53] T. F. O'Connor, M. E. Fach, R. Miller, S. E. Root, P. P. Mercier, and D. J. Lipomi. The language of glove: Wireless gesture decoder with low-power and stretchable hybrid electronics. *PloS one*, 12(7):e0179766, 2017. 2

[54] W. Piekarski and B. H. Thomas. Interactive augmented reality techniques for construction at a distance of 3d geometry. In *Proceedings of the workshop on Virtual environments 2003*, pp. 19–28, 2003. 2

[55] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pp. 39–ff, 1997. 5

[56] T. Piumsomboon, A. Clark, and M. Billinghurst. Physically-based interaction for tabletop augmented reality using a depth-sensing camera for environment mapping. *Proc. Image and Vision Computing New Zealand (IVCNZ-2011)*, pp. 161–166, 2011. 2

[57] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*, pp. 282–299. Springer, 2013. 2

[58] R. Poelman, O. Akman, S. Lukosch, and P. Jonker. As if being there: mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pp. 1267–1276, 2012. 2

[59] M. Prilla, M. Janßen, and T. Kunzendorff. How to interact with augmented reality head mounted devices in care work? a study comparing handheld touch (hands-on) and gesture (hands-free) interaction. *AIS Transactions on Human-Computer Interaction*, 11(3):157–178, 2019. 2

[60] M. Quandt, D. Hippert, T. Beinke, and M. Freitag. User-centered evaluation of the learning effects in the use of a 3d gesture control for a mobile location-based augmented reality solution for maintenance. In *DELbA@ EC-TEL*, 2020. 3

[61] R. Radkowski and C. Stritzke. Interactive hand gesture-based assembly for augmented reality applications. In *Proceedings of the 2012 International Conference on Advances in Computer-Human Interactions*, pp. 303–308. Citeseer, 2012. 2

[62] A. Robb, K. Kohm, and J. Porter. Experience matters: Longitudinal changes in sensitivity to rotational gains in virtual reality. *ACM Transactions on Applied Perception*, 19(4):1–18, 2022. 4

[63] N. Rosa. Player/avatar body relations in multimodal augmented reality games. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 550–553, 2016. 3

[64] B. Schiettecatte and J. Vanderdonckt. Audiocubes: a distributed cube tangible interface based on interaction range for sound design. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pp. 3–10, 2008. 2

[65] S. Schmidt, O. J. A. Nunez, and F. Steinicke. Blended agents: Manipulation of physical objects within mixed reality environments and beyond. In *Symposium on Spatial User Interaction*, pp. 1–10, 2019. 3

[66] R. Serrano, P. Morillo, S. Casas, and C. Cruz-Neira. An empirical evaluation of two natural hand interaction systems in augmented reality. *Multimedia Tools and Applications*, pp. 1–27, 2022. 3

[67] L. E. Sibert and R. J. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 281–288, 2000. 2

[68] T. A. Snijders and R. J. Bosker. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. sage, 2011. 6

[69] A. B. Soares, E. A. L. Júnior, A. de Oliveira Andrade, and A. Cardoso. Virtual and augmented reality: A new approach to aid users of myoelectric prostheses. *Computational Intelligence in Electromyography Analysis-A Perspective on Current Applications and Future Challenges*, pp. 409–426, 2012. 3

[70] M. Speicher, B. D. Hall, and M. Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019. 1

[71] G. Sroka, L. S. Feldman, M. C. Vassiliou, P. A. Kaneva, R. Fayez, and G. M. Fried. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial. *The American journal of surgery*, 199(1):115–120, 2010. 3, 5

[72] R. Terrier, F. Argelaguet, J.-M. Normand, and M. Marchal. Evaluation of ar inconsistencies on ar placement tasks: A vr simulation study. In *Virtual Reality and Augmented Reality: 15th EuroVR International Conference, EuroVR 2018, London, UK, October 22–23, 2018, Proceedings 15*, pp. 190–210. Springer, 2018. 8

[73] P. P. Valentini. Natural interface for interactive virtual assembly in augmented reality using leap motion controller. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 12(4):1157–1165, 2018. 2

[74] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. 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):2412–2422, 2023. 2, 3, 4, 5, 6, 8, 9

[75] 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, 2023. 6

[76] R. Y. Wang and J. Popović. Real-time hand-tracking with a color glove. *ACM transactions on graphics (TOG)*, 28(3):1–8, 2009. 2

[77] Wen. Wen 3921 16-inch two-direction variable speed scroll saw, 2023. https://wenproducts.com/products/wen-3921-16-inch-two-direction-variable-speed-scroll-saw. 3

[78] M. Whitlock, E. Harnner, J. R. Brubaker, S. Kane, and D. A. Szafir. Interacting with distant objects in augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 41–48. IEEE, 2018. 1, 2

[79] J. Wither and T. Hollerer. Evaluating techniques for interaction at a distance. In *Eighth International Symposium on Wearable Computers*, vol. 1, pp. 124–127. IEEE, 2004. 3

[80] E. Wolf, N. Döllinger, D. Mal, C. Wienrich, M. Botsch, and M. E. Latoschik. Body weight perception of females using photorealistic avatars in virtual and augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 462–473. IEEE, 2020. 3

[81] N. Yee, J. N. Bailenson, and N. Ducheneaut. The proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*, 36(2):285–312, 2009. 9

[82] A. Zenner and A. Krüger. Estimating detection thresholds for desktop-scale hand redirection in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 47–55. IEEE, 2019. 2

[83] H. Zhao and W. H. Warren. On-line and model-based approaches to the visual control of action. *Vision research*, 110:190–202, 2015. 8

[84] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 193–202. IEEE, 2008. 2