# Give Me a Hand: Improving the Effectiveness of Near-field Augmented Reality Interactions By Avatarizing Users' End Effectors

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

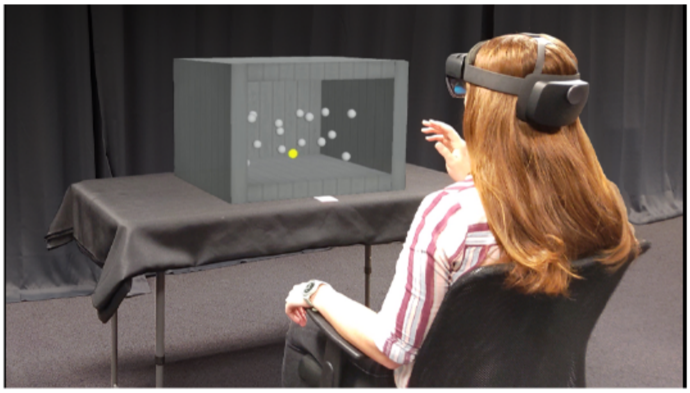


Fig. 1: Third person perspective of participant performing the object retrieval task. The holographic box is registered and augmented on top of the table.

**Abstract**—Inspired by previous works showing promise for AR self-avatarization - providing users with an augmented self avatar, we investigated whether avatarizing users' end-effectors (hands) improved their interaction performance on a near-field, obstacle avoidance, object retrieval task wherein users were tasked with retrieving a target object from a field of non-target obstacles for a number of trials. We employed a 3 (Augmented hand representation) X 2 (density of obstacles) X 2 (size of obstacles) X 2 (virtual light intensity) multi-factorial design, manipulating the presence/absence and anthropomorphic fidelity of augmented self-avatars overlaid on the user's real hands, as a between subjects factor across three experimental conditions: (1) No-Augmented Avatar (using only real hands); (2) Iconic-Augmented Avatar; (3) Realistic Augmented Avatar. Results indicated that self-avatarization improved interaction performance and was perceived as more usable regardless of the anthropomorphic fidelity of avatar. We also found that the virtual light intensity used in illuminating holograms affects how visible one's real hands are. Overall, our findings seem to indicate that interaction performance may improve when users are provided with a visual representation of the AR system's interacting layer in the form of an augmented self-avatar.

**Index Terms**—Interactions in AR, End-effector representation, Avatars, Augmented Reality

## 1 Introduction

Augmented reality (AR) is growing in popularity with technological conglomerates steadily invested in seeing this technology become ubiquitous. The technology allows to register computer generated two or even three-dimensional interactive virtual objects on to the real world [6]. This augmentation of virtual content onto the real world is commonly realized using either immersive optical see-through (OST) displays (e.g. Microsoft Hololens, Magic Leap, etc.) or video see-through (VST) displays (e.g. smartphones, tablets, Oculus quest, etc.). VST displays capture footage of the real world and apply in-painting techniques to add 3D content. OST displays on the other hand generate, render and display 3D content overlaid onto the real world. A number of domains such as those of medicine, architectural design, rehabilitation, and remote collaboration exist wherein augmentation of virtual content to the real world, can prove to be highly beneficial [1, 5, 25, 44, 69, 79, 80].

Interaction is central to AR (MR) applications wherein users interact with augmented 3D content projected onto the real world. These interactions can involve virtual entities like menus, interfaces, objects, virtual humans, etc. [6]. While interactions can involve virtual objects both in the near field as well as the far field, most AR systems involve near field interactions. These interactions can be supported through additional hardware like controllers or through natural means like speech, eye gaze, hand gestures, etc. [76]. Users tend to prefer natural interactions supported through free-hand gestures given their familiarity with them. For this reason, natural freehand gestures continue to be the de facto standard for interactions with virtual content in the near field.

In virtual reality (VR), users are usually provided with virtual representations termed user representations/self avatars [63]. These representations provide users with a frame of reference of themselves, visually depicting the locus of the user in the virtual scene. They also affect perceptions and the efficacy with which users can perform interactions

- Roshan and Rohith Venkatakrishnan are PhD candidates in the School of Computing at Clemson University. E-mail: rvenkat@g.clemson.edu, rohithv@g.clemson.edu
- Balagopal Raveendranath is a PhD student with the Department of Psychology at Clemson University. E-mail: braveen@g.clemson.edu
- Christopher C. Pagano is a faculty in the Department of Psychology at Clemson University. E-mail: cpagano@clemson.edu
- Wen-Chieh Lin is a faculty 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 faculty in the School of Computing at Clemson University. E-mail: arobb@clemson.edu, sbabu@clemson.edu

*Manuscript received 14 October 2022; revised 13 January 2023; accepted 30 January 2023. Date of publication 22 February 2023; date of current version 29 March 2023. Digital Object Identifier no. 10.1109/TVCG.2023.3247105*

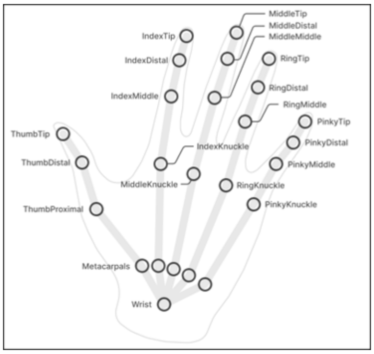



Fig. 2: Schematic representation of the joints and phalanges tracked by the MRTK framework on the Microsoft Hololens 2.

in virtual worlds [13, 36, 39, 43, 70], making them important characteristics of the virtual experience. A number of methods are used to provide self avatars in AR. These include using a holographic mirror to depict an avatar that the user embodies, overlaying virtual body parts over users' real bodies, and users viewing a third person perspective (3PP) of themselves depicted as an avatar they embody [19]. However, there is no explicit need to self-avatarize the user from an interaction standpoint, given that the real body continues to remain visible in AR. This being said, researchers continue to explore how self-avatarization can be beneficial from an embodiment perspective. Along these lines, avatars can be used to augment arms for users that may have missing limbs. This has many applications in the field of medical prosthetics. Muscular self avatars in AR have been shown to improve physical performance as a consequence of the Proteus effect [48]. Augmented arms have been shown to facilitate interactions with far field objects connected with an AR system without breaking users' sense of embodiment [17]. While such examples lend support for self-avatarization in AR, it remains to be seen if and how this affects interactions with virtual entities.

A number of AR applications used in the field of medicine, training, and surgery require users to perform near field fine motor interaction tasks that demand precise perception-action coordination. In scenarios like these, interaction performance is of paramount importance, and the research community hence continues to focus on improving the efficacy with which users perform interactions. On the one hand, provisioning users with a self avatar (overlaying virtual limbs on their physical counterparts) could confuse users because their real hands may still remain visible. On the other hand, avatarizing users may improve interactions given users' ability to embody and distalize to the augmented representation with which interactions are performed. This avatarization may effectively provide users with a visual representation of where the system approximates and tracks their limbs to be. Users may leverage this visualization to more accurately perform interactions given the apparent disparity between the actual position of a user's hand and its tracked position. With the majority of AR applications involving some form of near-field interaction with virtual content using tracked hand gestures, it is imperative to understand how augmented self-avatarization impacts the effectiveness of near-field interactions. Towards this end, we discuss the results of an experimental investigation that empirically evaluates how the presence and fidelity of an overlaid augmented self-avatar affects the performance of a fine motor object retrieval task. We discuss implications of self-avatarization for AR researchers to consider when designing interactive systems.

## 2 Related Works

### 2.1 Interaction in Augmented Reality

Interaction in AR usually involves users interacting with virtual entities like objects, interfaces, menus, etc., that are registered in 3D and superimposed in the real world [6]. These interactions can be supported through natural means like speech, eye gaze, hand gestures, facial expression or hardware driven methods that use hand-held controllers [76]. Researchers continue to extensively explore such modalities to support intuitive and immersive interactions with virtual entities in the near field [3, 82]. A majority of these research efforts suggest that users tend to be more efficient in performing near field interactions through direct manipulation, expressing preference for modalities that are supported through natural means like freehand gestures [9, 20, 53]. However, some work finds preferences for non-natural interaction systems over natural ones [55].

Hand-based AR interactions often require users to perform gestures like tapping a marker [11,45], making a fist or opening the palm [54,58], pinching or pushing with the finger tip [45, 52], or manipulating an object through space [14,57,61]. While some systems limit interactions to one hand [40, 58], other AR systems allow for two handed interactions. Usually, hand-gesture recognition in AR systems is realized either by the use of wearable data gloves [37, 49] or by using depth cameras, video cameras or infrared sensors [32, 65, 72, 73]. Despite wearables providing higher accuracy, reliability, and offering a potential for haptic feedback, gesture recognition through depth cameras, infrared sensors and other vision based systems tend to be preferred due to their simplicity, freedom of movement, and not requiring specialized hardware.

Interactions supported through natural hand gestures seem to be more suited for near field interactions than those involving manipulations of virtual objects that are situated at larger distances from the user. Along these lines, research shows that pointing to objects that are occluded may require nonlinear spatial and visual mapping in noisy environments [16]. Furthermore, far-field interactions may manipulate geometries that extend beyond the user's arm reach [22, 50]. Precision with respect to pointing and selecting also degrades with target size and distance to the target, making far field AR interactions challenging when supported through natural hand gestures [31]. For these reasons, multi-modal interaction modalities that combine technologies like speech, gesture recognition, etc., have been introduced and researched [27, 75]. In comparing free-hand gesture and multi-modal gesture-speech interactions, strengths and weaknesses have been discussed. Free-hand gestures tend to support better spatial input, while speech commands offer better system control [26, 34]. This can be understood with an example; a multi-modal interaction technology supporting gestures and speech allows users to gesture in order to identify the virtual object they want to interact with, and speak to perform some action on it [51]. Other research has found that speech tends to outperform gestures in terms of accuracy but the simplicity of gestures more than compensates for this loss in accuracy with speed [10]. It hence appears that the ambiguities and challenges associated with freehand gestures manifest predominantly when interactions involve virtual objects in the far field wherein an increased distance leads to the breakdown of direct manipulation metaphors [29].

The manipulation of virtual objects using hand gestures in AR can involve two methods, namely, metaphoric and isomorphic hand interactions [38]. The latter involves interaction systems that perform one-to-one literal spatial relations between input actions and resulting system effects whereas the former involves interactions based on image schemas and system effects on related conceptual metaphors. For example, a pinch gesture that allows a virtual cube to be manipulated through space would be a metaphoric approach whereas holding the cube by its edges involves an isomorphic interaction. Some research has shown that the isomorphic paradigm is perceived as more natural and usable when performing a displacement task whereas the metaphoric approach tends to be more appropriate for resizing tasks [18]. Other researchers have also found no differences between these interaction methods in terms of both task performance as well as users' subjective perceptions [66].

When AR interactions require precise manipulation of virtual objects, the modality used to support these interactions is highly influential. Researchers have hence studied scenarios in which certain input modalities support higher levels of performance in tasks in AR. Along these lines, it has been found that touch and freehand gestures are well suited for selection tasks involving individual virtual entities whereas voice commands excel for tasks that involve the creation of new visualizations [7]. For tasks that involve 3D cursor placement, it has been shown that users prefer handheld controllers with levels of performance being comparable when users used remotes and embodied head-tracked cursors [77]. A recent investigation comparing the effects of different input modalities on a Fitts' law based target selection task in AR demonstrated that the opacity or rather transparency level of the target has little to no effect on performance. The aforementioned study, however, suggests that a ray-cast based selection technique outperforms both a touchpad and gesture based approach in terms of throughput and error rates [41]. This being said, near field interactions in AR that involve selection and manipulation of virtual objects often continue to leverage hand gestures as the means of realizing user interactions seeing as how users are familiar with this method of interaction, learning how to gesturally control virtual objects in a short amount of time [56].

### 2.2 Effects of Avatars on AR Interactions

Avatarizing users in AR is gaining popularity given the potential for users to perceive embodiment like in VR. This opens the door to a numerous applications in fields such as medical practices [33], education [28], remote collaboration [47], video games [60], etc. Avatarization in AR can be achieved in a number of ways based on the display device (e.g. head mounted, handheld) used, the rendering technique (e.g. optical, video) adopted, and the user perspective leveraged (1PP and 3PP). OST HMDs overlay augmented avatar content on the real world while VST HMDs refer to VR headsets equipped with external cameras that combine live image processing with in-painting techniques to modify (erase, embellish, accessorize) content captured via the cameras to display avatars. Holographic augmented mirrors can be augmented to display an avatar that the user embodies and this technique has been shown to influence body weight perception [78]. Similar to the mixed reality continuum [42], the degree of avatarization in AR follows a continuum that ranges from containing no virtual elements (no-avatar or real body) on one end to full avatarization on the other, wherein users embody and control a complete virtual body [19]. Just before full avatarization is a region corresponding to partial avatarization wherein human limbs are replaced or overlaid with virtual counterparts. This avatarization finds high relevance in medical fields like prosthesis for severed patients [68], and rehabilitation and recovery [21, 23, 30, 59]. A recent survey article summarizes literature that pertains to embodiment through avatarization in AR [19].

Research on manipulating the appearance of the users' self avatar in AR shows that when embodying a more muscular avatar, users physical performance improves [48]. Recent work that investigated the effects of hand representations in AR has shown a feasibility of using AR avatar arms that are expandable to interact with far field real word objects connected to the system [17]. Users' virtual arms were made twice the length of the users' real hands and it was found that interaction with far field objects was possible without breaking the users' sense of embodiment. However, the real world's visibility was found to be a hindrance to the sense of embodiment. While such efforts study how avatar representations affect perceptions and interactions in AR settings, it continues to remain a relatively unexplored field. Most research that involves avatarization of the user in AR tends to focus on the sense embodiment without getting into the depths of the mechanics of near-field fine-motor interactions in AR. It follows that the effects of avatars and their characteristics on interactions also remains unexplored. It is hence favourable to draw from related literature on VR research focusing on avatars and interaction when choosing experimental tasks and manipulations.

Users in VR are often provided with virtual representations that are commonly referred to as user representations [63, 64]. These representations help users perform actions and affect how effectively, accurately, and efficiently users can perform them [13, 39, 43, 70]. User representations can differ in terms of their visual appearance, the input modality used to support their actions, and the mapping of their inputs to control mechanisms [64]. End-effector representations can be considered a subset of user representations which involve the virtual representations of the hands or the tools with which actions are performed in VR. These end-effectors refer to the virtual tool or the part of the representation that comes into contact with the object being manipulated [2]. Similar to user representations, the way in which end-effectors are represented affect task performance, embodiment and other characteristics associated with interactions in IVEs. In terms of task performance, research suggests that simplified end-effector representations produce faster and more accurate interactions for a pick and place task [4]. In a study comparing three representations of virtual arms (i.e., 'hand-only', 'hand+forearm' and 'whole arm") on the performance of a collision-avoidance based selection task, it was found that that representing the end-effectors as whole arms made users take more time to perform the selection than when representing the end-effectors as just hands [71]. In the aforementioned study, users were tasked with selecting randomly assigned targets one after another from an interaction volume that hosted a field of non-target spheres. The number (density) and size of non-target obstacles within the volume were manipulated and studied. We draw inspiration from this task given the precise fine motor perception-action coordination required in performing it.

### 2.3 Impact of Real Hand Visibility

Unlike immersive VR, using an AR display does not necessarily obscure users from being able to see their real body, especially when using an OST display or a projection based system. The colors and light intensities applied on virtual holograms affects how visible one's real hand is when interacting with the holograms. Research on this front has measured the strength of the virtual hand illusion under different virtual:real light intensity ratios. It was found that ratios of 0.75:0.25 and 1:0 produced optimal results on a typing task without compromising the level of ownership or agency towards the virtual hand for users that had trembling hands [74]. These results suggest that real body visibility need not necessarily impede embodiment experiences from a 1PP as long as the virtual hands are more visible than the users real hands [19]. In terms of the transparency of the real hand (real hand visibility), users prefer interacting when their hands are more visible with lower transparency (alpha of 0.6 and 0.8) than when their hands are more transparent (alpha of 0.5 or lower) [8]. Users in this study found it odd and absurd to interact when their hands were less visible, commenting that it was alien-like and weird. Furthermore, it was observed that the perceived level of transparency varied depending on the background that the hand was being viewed against. However, it must be noted that this study did not afford users with an augmented self-avatar. In general, it is observed that increasing the virtual light intensity is associated with decreased real body visibility. Overall, the impact of real hand visibility on interaction performance in AR is an area that warrants more concrete investigations.

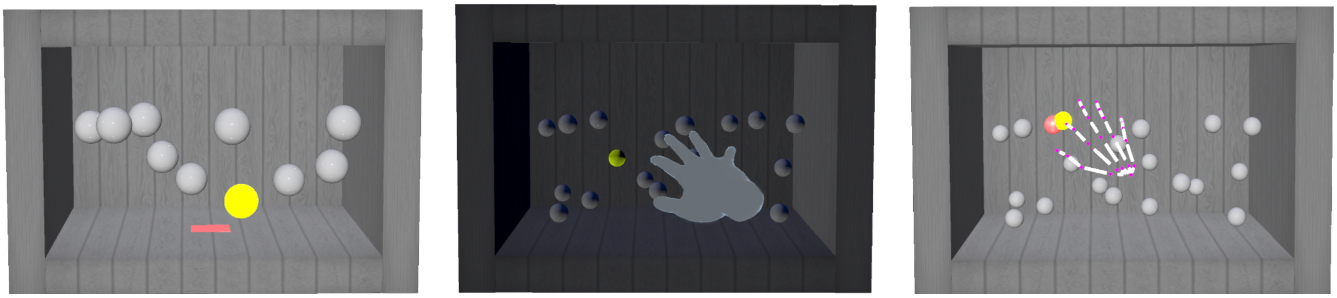
## 3 System Description

### 3.1 Apparatus

The virtual holograms used for this study were built using the Unity 2020.2.2f1 game engine on a computer equipped with an Intel i7-8700 processor, 32 GB of RAM, and an NVIDIA RTX 2080 graphics card. The application was deployed on a Microsoft Hololens 2 HMD. The Hololens 2 has a 54° diagonal FOV with a frame refresh rate of 60 Hz. Users were seated on a rolling chair in front of a physical table atop which sat the holographic box and its contents. The table was covered with a black cloth to facilitate better viewing experiences of the augmented holograms.

### 3.2 Virtual Components (Holograms)

A virtual wooden box (hologram) was used to host the object retrieval task described in section 4.1. The box was modeled to be 47cm wide, 45cm deep, and 33.65cm high, and was spacious enough to host 20 balls interspersed within its volume. The front of the box was left open and allowed for users to reach in and retrieve the designated target on each trial of the experiment. A wooden texture was applied on the walls of the box to make the target and non-target balls appear more salient. Textures applied to the balls (both targets and non-target obstacles) were chosen such that their contours were clearly visible and suggestive of their boundaries. In performing the trials, auditory feedback of the different events (selection, collision with other non-target obstacles, collisions with surfaces of the box, trial completion) were fed to the



(a) No Augmented Avatar. (b) Realistic Augmented Avatar. (c) Iconic Augmented Avatar.

Fig. 3: Holographic content augmented in the three conditions of the study. Each sub-figure depicts the virtual holographic box containing the target to be retrieved (colored in yellow) surrounded by non-target obstacles (colored in grey). Sub-figure (a) depicts a surface collision during a trial featuring the low density and large obstacle-size pair configuration, while (b) shows a low virtual light intensity for a high obstacle density, small size pair configuration. Subfigure (c) shows a target obstacle collision during object retrieval.

participant using the built-in speakers providing a head related transfer function (HRTF), allowing to simulate more accurate spatial sounds.

The box was registered on top of the physical table using a customized script that allowed to adjust its perceived position and orientation on the table. A physical marker affixed to the table was used as a positional and orientational reference to consistently register the box at the same position on the table. At the start of the experiment, the registration of the box was carried out to establish a perfect co-location between the virtual 'start trial' button (section 4.1) and the physical marker. Furthermore, the positions of other physical objects in the rest of the room remained unchanged in order to ensure that both the spatial mapping detected by the headset as well as the real world background remained constant across all participants. These measures allowed us to ensure that the virtual box was consistently placed at the same position with the same orientation on the physical table for each participant, ensuring consistency in spatial manifestation of the holograms both between and within conditions. Overall, for all users across the conditions, the holographic box along with the button was augmented in the same position on the table, with the background, and lighting of the real world remaining unchanged.

### 3.3 Hand Representations

An augmented hand avatar involves the provision of a tracked avatar (hologram) based on where the headset detects and tracks the user's real hand using camera vision. Two augmented avatar representations were compared to a baseline condition without any augmented avatar, making a total of three experimental conditions (figure 3):

**No Augmented Avatar:** In this condition, participants interact with holograms without the provision of any augmented (visualized) avatar hands. Given that the technology investigated involves OST display based AR, users can see their own real hands using which they can perform interactions.

**Iconic Augmented Avatar:** In this condition, participants interact with holograms with an augmented (visualized) avatar hand. This avatar is represented as a combination of joints and bones and resembles an iconic representation of a hand. The joints are colored pink and the bones are colored white to make the augmented avatar salient. The augmented avatar is animated to move in real time based on the users own hand movements tracked by the AR system. In order to realize this augmented avatar representation, a custom hand visualization script was conceptualized and developed. The script is programmed such that for each frame that hands are detected by the built-in cameras in the Hololens 2, the positions and orientations of each of the 25 joints detected by the mixed reality toolkit (MRTK) framework are obtained. A small pink sphere is then rendered at the position of each tracked joint for that frame. The script also renders capsules to represent bones/phalanges between every two successive joints corresponding to the same finger, on every frame. In rendering the joints and bones for all the fingers in every frame that the hand is tracked by the system, a fully tracked augmented avatar hand is realized that is animated to move based on the users' real hands in real time.

**Realistic Augmented Avatar:** In this condition, participants interact with holograms with an augmented (visualized) avatar hand. This avatar is represented as a human like hand wearing a white glove. The augmented avatar is animated to move in real time based on the users own hand movements tracked by the AR system. A hand visualizer script provided by the MRTK framework was used to control a skinned mesh to visualize the tracked hands based on the positions and orientations of the tracked joints and bones of the hand. This works when the hands are detected by the cameras built into the Microsoft Hololens 2. The texture and shaders applied on the virtual hand made the representation appear neutral as if the user was wearing a white glove. This was done in order to prevent any feelings of eeriness that could arise from choosing realistic human textures that are not matched to the user in terms of color and gender, an effect that prior research has demonstrated [62]. We consciously avoided using realistic human skin tones. Furthermore, this made the avatar augmentation appear salient rather than a blend between the users actual and augmented hands.

In all conditions, the interactions of users' hands with the holograms of the balls and box are based on where the HMD tracks the hands to be. The two avatarization conditions (iconic and realistic) feature an augmented visualization that visually represents the HMD's tracking of the hands. The latter involves a rigged mesh of higher anthropomorphic fidelity while the former involves an iconic representation of an augmented virtual hand.

A system evaluation was conducted to calculate the positional offset between the augmented avatars and users' real hand position. The distance between the tip of the index finger of the real hand and the avatar's fingertip was measured over 10 samples using a ruler while resting the physical hand on top of the table. This average computed offset was found to be(M=3.66mm, SD=0.69). A system evaluation of latency and frame rate in all three conditions was conducted using Niehorster et al.'s method [46]. 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-Avatar (Pos. lag =28.33ms and Ori. lag = 28.75ms), Iconic avatar(Pos. lag =29.58ms, Ori. lag=28.75ms), Realistic avatar (Pos.lag =29.58ms, Ori. lag = 29.16ms).

### 3.4 Virtual Light Intensity

The intensity of virtual light (applied to virtual components) in optical see-through AR can affect the perceived visibility of the real body [74]. With this in mind, we investigated two different levels of virtual light intensity, aiming to study how the resultant visibility of one's real hand impacts interactions in the presence and absence of augmented avatars. The two levels of the light intensity investigated are termed as high and low intensity. For the low virtual light intensity, two virtual directional light objects were added in the Unity scene to illuminate the box and its contents (balls). These were oriented in the front-back and left-right direction. For the high virtual light intensity, an additional three virtual directional light objects (front-back) were added to the scene. The indirect multiplier and intensity parameters of all the virtual lights were set to 1, and the cucoloris mask parameters of all lights were set to 10. The color of all the virtual lights were set to white (hex:#FFFFFF) to match the color of the lighting in the room. The intensity of light in the real world was measured using a procured light meter [1], and it was ensured that the lighting conditions in the room was kept constant for all participants throughout the course of this study.

## 4 Experiment

### 4.1 Task

For this experiment, a collision avoidance based object-retrieval task was conceptualized wherein users had to retrieve an augmented virtual (holographic) target ball from a field of obstacles (non-target balls) within the virtual box for a number of trials (figure 1). The box was augmented atop a real world physical table and had a number of identical non target balls that were scattered within its volume. Users were seated on a chair in front of the table and were tasked with retrieving the target ball while avoiding collisions with other non-target balls and the box. The front of the box was left open for users to be able to see the interaction volume (the space inside the box) within which the balls were scattered. A similar task was used in a study investigating the effects of virtual arm representations on selection interactions in virtual reality [71].

At the start of every trial, the box (interaction volume) was empty. To initiate each trial, users had to push down a virtual button augmented in front of the box. When pressed, the button would disappear and the crowd of balls from which the target had to be retrieved would appear inside the box, marking the start of that trial. This ensured that all participants began every trial with their hands in the same position. In each trial, the target to be retrieved was rendered in yellow while obstacle balls were rendered in gray, clearly delineating the target from the deterrent non targets. A touch and snap interaction metaphor was utilized as the means for users to select and maneuver the target through the interaction volume. Accordingly, when a user touched the target with the tip of their dominant hand's index finger, the target would snap to their fingertip, generating a selection sound deployed via the headset. Upon this selection, the user could carefully retract their hand along with the snapped target, maneuvering it through the interaction volume. Users were tasked with retrieving the target from inside the box while avoiding collisions with any of the obstacles which included both the non-target balls inside the box and the box in itself.

During a trial, if either the user's tracked hand or the target to be retrieved collided with any of the other non-targets or the surfaces (walls, floor and ceiling) of the box, feedback of the collisions were provided both visually as well as aurally. The visual feedback provided during a collision with a non-target ball involved that non-target ball being highlighted in red for as long as the collision was taking place. Similarly, collisions with any of the surfaces of the box involved the center of that surface turning red for as long as the collision was taking place. The auditory feedback provided for a collision with a non-target ball involved a collision sound being played to the user via the headset. Similarly, a surface collision sound was played via the headset whenever collisions occurred with any of the box's surfaces. When users managed to successfully retrieve the ball from inside the box, bringing it completely outside the interaction volume, the target and obstacles would immediately disappear and users were provided with a success sound, giving them both visual and auditory feedback that was indicative of the completion of a trial. The completion of a trial was further marked by the reappearance of the virtual button that had to be pressed to initiate the next trial. It was ensured that the sounds associated with target selection, trial completion, non-target collisions, and surface collisions were distinct and different from each other, and that the visual and auditory feedback occurred simultaneously thus providing users with multi-modal feedback that was indicative of the events that occurred during a trial. Users were free to take as much time as they required for each trial in order to minimize the number of collisions. Thus, they were instructed to focus more on avoiding collisions rather than completing trials quickly.

[1] https://drmeter.com/products/digital-lux-meter-digital-illuminance-light-meter-lx1010b-dr-meter

### 4.2 Study Design

To empirically evaluate how the presence and anthropomorphic fidelity of virtual hand representations affects users' performance in a near field AR interaction task, we employed a 3 (augmented hand representation) X 2 (density of obstacles) X 2 (obstacle size) X 2 (virtual light intensity) multi-factorial design, manipulating the virtual end-effector representation as a between-subjects factor across three experimental conditions: (1) No Augmented Avatar (no hologram overlaid); (2) Iconic Augmented Avatar (overlaying an avatar hologram represented as tracked joints and bones); (3) Realistic Augmented Avatar (overlaying an end-effector hologram represented as a tracked hand). Users in each condition performed an object retrieval task for a number of trials in which the density of the obstacles (number of balls within interaction volume), the size of obstacle balls, and the virtual light intensity (which in turn affects users visibility of their real hands) were manipulated as within subjects factors.

For each experimental condition, participants performed the object retrieval task described in section 4.1 for a total of 160 trials. Two different obstacle density configurations categorized as low and high were tested in this study. The lower density configuration consisted of ten identical balls scattered within the interaction volume and its higher counterpart included an additional ten balls(identical) making a total of 20 balls. Thus, the higher density configuration involved a more crowded interaction volume with the same space hosting twice the number of balls, half of which were in the same positions as the balls in the lower density configuration. For each density configuration, two different sizes of balls categorized as small (diameter=2.9cm) and large (diameter=5.5cm) were utilized. For a given density configuration, the positions of the centers of all balls remained the same regardless of the size of the balls. For any given trial, all balls (both the target and the non-target obstacles) were identical in size. Participants performed 20 retrievals (trials) for each of the four density-size pair combinations; once for each of the 20 balls in the high density configuration, and twice for each of the 10 balls in the low density configuration. Given two density configurations and two ball sizes, this accrued up to a total of 80 (20 retrievals * 2 density configurations * 2 ball sizes) trials, the order of which was randomized. Participants performed a second block of 80 randomized trials where the intensity of the virtual light was varied between these blocks. Two different virtual light intensity levels were investigated in this study, namely low and high intensity. The order of the light intensity blocks was counterbalanced to prevent extraneous influences of order. Thus, a participant was tasked with performing two blocks of 80 trials (160 trials) with each trial randomly featuring one of four density-size pair combinations along with a randomly chosen target specific to that combination.

### 4.3 Measures

**Performance** - Given that participants were instructed to prioritize minimizing collisions rather than completing the trials quickly, performance was measured as a function of effectiveness rather than completion time. The total number of collision events that occurred in each trial was used as an operational measure of performance. This was computed by incrementing a counter whenever any tracked joint or phalange of the user's hand collided with any of the non-target balls or surfaces of the box during that trial. The joints and phalanges considered for collision are depicted in figure 2. Hypothetically, if a user's entire hand were to pass through (collide with) a non-target ball, the total collision events recorded as a result would be equal to the number of joints plus the number of phalanges (25 joints + 24 phalanges = 49) in the user's tracked hand. Essentially, a higher number of collision events corresponds to worse performance.

**Perceived Real Hand Visibility** - At the end of each block, users reported how visible they perceived their actual hands (excluding forearm and wrist) to be on a 10 point scale. This scale was anchored such that a value of 10 on this scale corresponded to a user being able to fully see their real hand. Similarly, a value of 1 on this scale meant that users were not able to see their real hand.

**Perceived Usability** - Users' perceived level of usability associated with performing interactions with virtual holograms was measured using the PSSUQ (Post-Study System Usability Questionnaire), that originates from the IBM Usability scale. The PSSUQ is a sixteen item standardized questionnaire that is widely used to measure usability and perceived satisfaction. Counterintuitvely, a lower score on this scale corresponds to a higher perceived usability [35].

### 4.4 Research Question and Hypotheses

The overarching aim of this work was to answer the following research question: **how does augmented self-avatarization of users' hands (end-effectors) affect interaction performance in a near-field, fine-motor, obstacle-avoidance based object retrieval task in OST AR?** Downstream of this, we were interested in understanding how interactions based on such representations are affected by virtual light intensity. We operationalize user performance based on the measure described in section 4.3. We developed the following hypotheses that reflect work discussed in section 2:

**H1**: Users with augmented avatar representations will perform better than those without an avatar.

**H2**: A higher obstacle density configuration will be associated with worse performance.

**H3**: The size of the obstacles will affect performance.

**H4**: Performance will be worse with a higher virtual light intensity.

**H5**: Deeper targets will result in poorer performance.

It is expected that provisioning users with an augmented self avatar will give them a visual representation of the interacting layer (tracked position of their hands). This will allow them to distalize to the task dependent virtual end-effector rather than having to rely on their hands, potentially resulting in better performance. With respect to obstacle density, the high density configuration is likely to result in inferior performance due to the increased potential for collisions when then are more obstacles within the interaction volume. With target size, two competing expositions can be offered. On the one hand, a larger target can be reached more easily by virtue of its size, thus making its retrieval easier. On the other hand, the larger target is more likely to collide with other obstacles during retrieval than a smaller one. For these reasons, we do not develop directional hypotheses with respect to how the target size affects performance. An increased virtual light intensity is expected to obscure the visibility of users' actual (real) hands which in turn may worsen performance. With regards to the target depth, the number of obstacles that need to be avoided is larger for deeper targets, consequently resulting in a larger number of collisions.

### 4.5 Participants

A total of 54 participants were recruited for this Institutional Review Board (IRB) approved study, with 18 allotted per condition. Data from three participants (one from each condition) was excluded from analysis due to data logging errors. This led to a total of 8160 trials of object retrievals for analysis. Participant ages ranged from 18 to 30 years old (M=20.92, SD=3.96); 28, 21 and 2 of whom identified as female, male, and non-conforming respectively. All participants had normal/corrected-to-normal (20/20) vision. Overall, VR and AR experience did not significantly differ across conditions.

### 4.6 Procedure

Upon arrival to the laboratory, participants were greeted and asked to read and sign a consent form (informed consent). After consenting to participate in the study, participants filled out a demographics questionnaire that included information about their backgrounds and experience with AR, VR, video games, experience playing sports, etc. Following this, participants' arm lengths, interpupillary distances (IPD), and stereo acuity were measured. They were then randomly assigned to one of the three experimental conditions. The experimenter then detailed the task they would be performing in the study, demonstrating how to perform the gestures required for the object retrieval task (see section 4.1). Participants then donned the Microsoft Hololens 2 OST HMD after which an eye calibration routine was run to customize the experience, allowing for optimal hologram viewing and interaction. Participants then performed 5 practice trials to familiarize themselves with the task and its mechanics. It was ensured that the density and size of the balls presented in the practice phase were different from those tested in the experiment, thus preventing participants from learning the specific ball positions and sizes used in the trials of the experiment. After the practice phase, participants then began the experiment performing the task over the 160 trials. They were allowed to take a break whenever desired, but it was ensured that breaks were taken prior to the start of any trial. Upon completion, participants removed the HMD and filled out the PSSUQ [35]. They then proceeded to engage in a short semi-structured interview with the experimenter to discuss their experience in this study, the strategies they used, and aspects they found challenging about the task. They were then debriefed about the study and were compensated for their time. On average, it took a participant up to 50 minutes to complete the whole procedure.

## 5 Results

Since a repeated measures design was used in this experiment, variables had considerable nesting. As each participant completed 160 trials, a portion of the variance in their responses can be attributed to a common source – the fact that the same participant was responding to each trial. Level 1 (within-participant) variables represent those that change from trial to trial. Level 2 (between-participant) variables represent those that change from participant to participant. To properly account for variance between and within subjects, Hierarchical Linear Modeling was used [24]. Since the dependent variable is a count variable (total number of collisions events while retrieving the ball), performing a poisson regression is ideal in this case. However, while performing poisson regression, if the dispersion parameter $\phi$ is greater than 1 (overdispersion), a negative binomial regression can be used [12]. In this data, since $\phi$ was found to be 3.45 for the model with only the fixed effects, negative binomial regression was used to fix overdispersion.

For each analysis, 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 interactions, individual interaction terms were added to the main effects model one at a time. In each iteration of the model, there was never more than one interaction term present at a time. Results of each interaction are reported from the model in which that interaction was included. Effect sizes 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 [67]. Prior to conducting analysis, the extent of nesting in the data was assessed by computing the intraclass correlation coefficient (ICC) from the null model. The ICC was calculated to be 0.416, indicating that approximately 41.6% of the variance in the total number of collisions was associated with the participant and that the assumption of independence was violated. Following a multilevel modeling technique is ideal in this case. For all the following models, the only random effect computed was the intercept based on the Participant ID.

A negative binomial regression was conducted to assess the effects of condition, obstacle density, obstacle size, virtual light intensity and target depth on total collisions. This model with only the main effects (AIC = 27426, df = 9) offered a significantly better fit to the data than did the null model (AIC = 28945, df = 3), $\chi^2$ = 1531, $p < 0.001$. It explained 65.6% of the variance in total collisions (conditional $R^2$ = 0.66, marginal $R^2$ = 0.57).

### 5.1 Performance

#### 5.1.1 Augmented Hand Representation

A significant effect of avatar condition on total collisions was found, $\chi^2$ (2, N = 8160) = 154.36, $p < 0.001$, $sr^2$ = 0.29. Total collisions were significantly more in the No-Avatar condition (M = 2.08, SE = 0.17) as compared to the Iconic (M = -0.69, SE = 0.17), z = 11.48, $p < 0.001$,

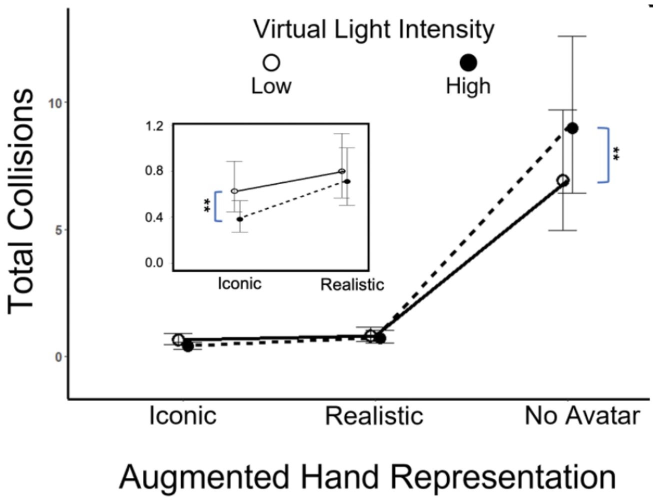



Fig. 4: Interaction between the avatar condition and real hand visibility. Error bars indicate 95% confidence intervals.

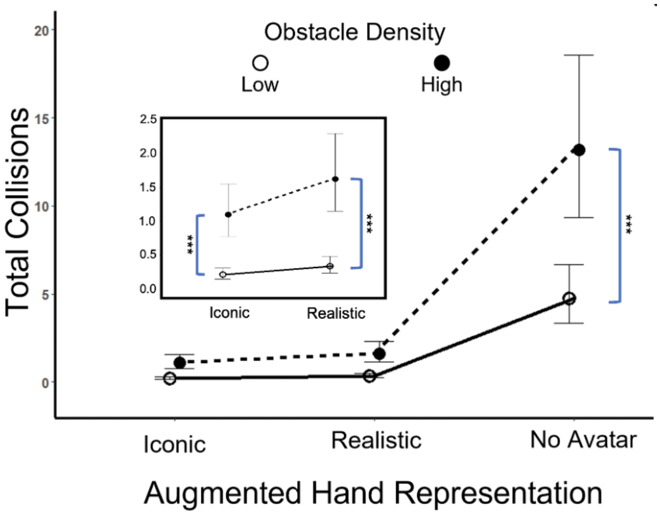



Fig. 5: . Interaction between the avatar condition and obstacle density. Error bars indicate 95% confidence intervals.

as well as the Realistic avatar condition (M = -0.28, SE = 0.17), z = 9.81, $p < 0.001$. There was no significant difference in total collisions between the Iconic and Realistic avatar conditions.

### 5.1.2 Density, Size, Depth and Virtual Light Intensity

There was a significant effect of obstacle density on total collisions, $\chi^2$ (1, N = 8160) = 630.25, $p < 0.001$, $sr^2 = 0.08$. Total collisions were significantly more for the high obstacle density (M = 1.05, SE = 0.10) as compared to the low density (M = -0.31, SE = 0.10).

Obstacle size also had a significant effect on total collisions, $\chi^2$ (1, N = 8160) = 20.02, $p < 0.001$, $sr^2 = 0.001$. Total collisions were significantly more for large obstacles (M = 0.49, SE = 0.10) as compared to small obstacles (M = 0.25, SE = 0.10).

The distance to the target also significantly affected the total collisions, $\chi^2$ (1, N = 8160) = 1238.25, $p < 0.001$, $sr^2 = 0.13$. For every one unit increase in distance to target, the difference in the log of expected count of total collisions is expected to change by 10.78 units. There were no main effects of virtual light intensity.

### 5.1.3 Interaction Effects

There was a significant interaction between avatar condition and virtual light intensity, $\chi^2$ (2, N = 8160) = 34.14, $p < 0.001$, $sr^2 = 0.005$ (figure 4). When testing simple effects, for participants in the Iconic avatar condition, total collisions were significantly more for the low virtual light intensity (M = 0.63, SE = 0.11) as compared to the high virtual light intensity (M = 0.38, SE = 0.07), z (2719) = -2.80, p = 0.005. For participants in the No-Avatar condition, total collisions were significantly more for the high virtual light intensity (M = 9, SE = 1.55) as compared to the low virtual light intensity (M = 6.92, SE = 1.20), z (2719) = 2.87, p = 0.004. However, for participants in the Realistic avatar condition, total collisions were not different between the virtual light intensities.

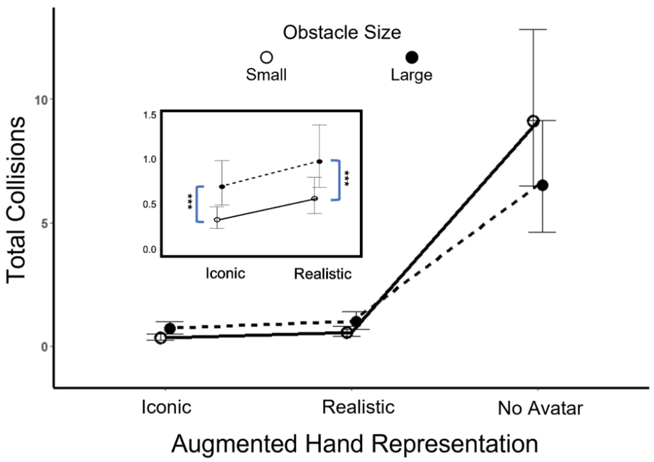



Fig. 6: Interaction between the avatar condition and obstacle obstacle size. Error bars indicate 95% confidence intervals.

There was a significant interaction between avatar condition and obstacle density, $\chi^2$ (2, N = 8160) = 27.34, $p < 0.001$, $sr^2 = 0.005$ (figure 5). When testing simple effects, for participants in the Iconic avatar condition, total collisions were significantly more for the high obstacle density (M = 1.09, SE = 0.19) as compared to the low obstacle density (M = 0.22, SE = 0.04), z (2719) = 12.49, $p < 0.001$. Similarly, when participants were in the Realistic avatar condition, total collisions were significantly more for the high obstacle density (M = 1.62, SE = 0.29) as compared to the low obstacle density (M = 0.34, SE = 0.06), z (2719) = 11.23, $p < 0.001$. Also, when participants were in the No-Avatar condition, total collisions were significantly more for the high obstacle density (M = 13.18, SE = 2.30) as compared to the low obstacle density (M = 4.73, SE = 0.83), z (2719) = 10.83, $p < 0.001$. The interaction effect observed is due to the larger difference between the low and high obstacle densities in the No-Avatar condition.

There was a significant interaction between avatar condition and obstacle size, $\chi^2$ (2, N = 8160) = 84.01, $p < 0.001$, $sr^2 = 0.008$ (figure 6). When testing simple effects, for participants in the Iconic avatar condition, total collisions were significantly more for the large obstacle size (M = 0.69, SE = 0.12) as compared to the small obstacle size (M = 0.33, SE = 0.06), z (2719) = 7.73, $p < 0.001$. Similarly, for participants in the Realistic avatar condition, total collisions were significantly more for the large obstacle size (M = 0.97, SE = 0.17) as compared to the small obstacle size (M = 0.56, SE = 0.10), z (2719) = 5.86, $p < 0.001$. However, for participants in the No-Avatar condition, total collisions were not significantly different based on obstacle size.

## 5.2 Perceived Real Hand Visibility

A Mixed model ANOVA was carried out to analyze the effect of condition and virtual light intensity on users' perceived real hand visibility. At the between-participants level, a significant main effect of condition was found, F(2,48) = 13.93, $p < 0.001$, ${\eta_p}^2 = 0.367$. Post-hoc pairwise comparisons using Tukey's HSD revealed that the perceived visibility of the hand was significantly higher in the No-Avatar condition (M = 6, SE = 0.378) when compared to both the Iconic (M = 3.82, SE = 0.37, $p = 0.001$), and Realistic (M = 3.35, SE = 0.378, $p < 0.001$) avatar conditions. At the within-participants level, a significant main effect of virtual light intensity was found F(1,48) = 93.117, $p < 0.001$, ${\eta_p}^2 = 0.660$. Perceived visibility was significantly higher when the virtual light intensity was low (M = 5.61, SE = 0.241) as compared to when the virtual light intensity was high (M = 3.17, SE = 0.263, $p < 0.001$).

On examining how condition moderates the effect of virtual light intensity on users' perceived real hand visibility, a significant interac-

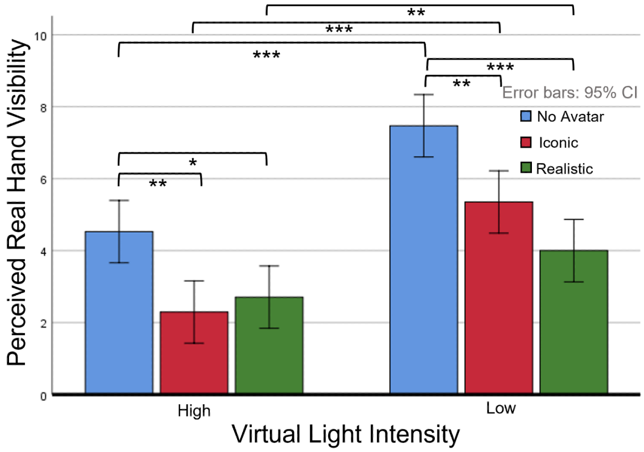



Fig. 7: Effect of virtual light intensity on perceived real hand visibility moderated by condition.

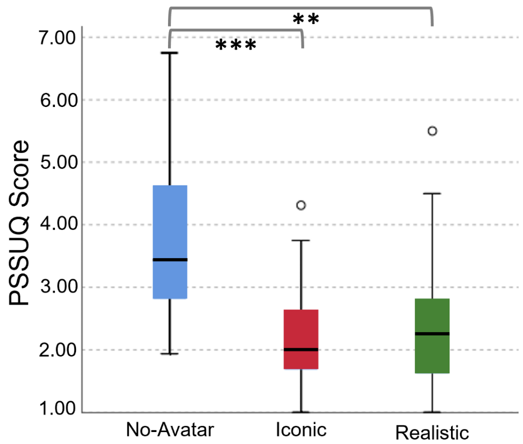



Fig. 8: Effect of condition on perceived usability. A lower PSSUQ corresponds to a higher perceived usability

tion effect was found, F(2,48) = 5.11, p = 0.01, $\eta_p{}^2 = 0.176$. When the virtual light intensity was high, users in the No-Avatar condition (M = 4.53, SD = 2.42) perceived their real hands to be significantly more visible than users in the Iconic (M = 2.29, SD = 1.04, p = 0.003), and the Realistic (M = 2.71, SD = 1.89, p = 0.02) avatar conditions. When the virtual light intensity was low, users in the No-Avatar condition (M = 7.47, SD = 1.84) perceived their real hands to be significantly more visible than users in the Iconic (M = 5.35, SD = 1.83, p = 0.002), and Realistic (M = 4.0, SD = 1.45, $p < 0.001$) avatar conditions. Inspecting this interaction effect by looking at the effect of virtual light intensity across conditions, perceived real hand visibility was significantly higher for the low virtual light intensity compared to the high virtual light intensity in all conditions. However this difference was less pronounced for the realistic avatar as evinced in figure 7.

### 5.3 Post Study System Usability

A Kruskal-Wallis H test showed that there was a statistically significant difference in the perceived usability score between the different conditions, $\chi^2(2) = 16.811$, $p < 0.001$. Post-hoc pairwise Mann Whitney tests using a Bonferroni-adjusted alpha level of 0.17 (0.05/3) were used to compare every pair of conditions. A significant difference in usability score was found between the No-Avatar and Iconic avatar conditions, U($N_{No-Avatar}$ = 17, $N_{Iconic}$ = 17) = 34.0, $z = -3.81$, $p < 0.001$; and between the No-Avatar and Realistic avatar conditions U($N_{No-Avatar}$ = 17, $N_{Realistic}$ = 17) = 50.5, $z = -3.24$, $p = 0.001$. Participants in the No-Avatar condition (Mdn = 3.43) perceived less overall usability than participants in the Iconic (Mdn = 2.00), and Realistic (Mdn = 2.25) avatar conditions. Figure 8 depicts these results.

## 6 Discussion

The statistical analyses pertaining to users' interaction performance revealed that the presence of an augmented avatar hand representation significantly impacted users' ability to perform the target retrieval task. The results indicated that users performed significantly better when they were afforded an augmented avatar compared to when they were not provided with one. This offered support for hypothesis H1, suggesting merit in avatarizing users' end-effectors for near field interactions that involve precise motor control and perception-action coordination. The main factor affecting how well users perform this task revolves around the visibility of the interacting layer (i.e. where the system senses and tracks the user's hand to be). Interactions in any AR system are based on this layer, and some researchers refer to this layer as the tracking module of the system. It is reasonable to expect that any AR system would have a disparity between the position of the user's actual hand and the interacting layer (tracked position) [19]. Both the avatar conditions provide users with a visual representation of this interacting layer. Thus by avatarizing the user with an augmented end-effector, they can perform the task ensuring that their avatar does not collide with any of the obstacles or the box. The provision of an avatar thus allows users to distalize to a virtual representation of the tracked hand based on which interactions actually take place. Without an avatar, however, users have to perform the task based on their real hands attempting to infer where the interacting layer is (erroneously assuming that it is where their real hand is) rather than distalizing to a visualization of it. Even if the visual feedback obtained during collisions could have indirectly provided information about the interacting layer, it is far from a direct visualization of the layer. Users in the No-Avatar condition hence would have to rely on proprioceptive information about where their hands were in space, in addition to the visual information of where their actual hands were, to approximate the location of the interacting layer. The absence of a direct visual representation of the interacting layer hence explains why performance was worse in this condition. This suggests that abstracting the user from the interaction layer can cause detriments to performance in tasks that involve precise perception-action coordination.

With respect to the factors affecting interaction performance at a within-participants level, our results showed significant effects of the obstacle density (number of obstacles within the interaction volume), the size of the obstacles, depth of the target. Firstly, as expected, participants performed worse when retrieving targets from a higher density of obstacles than when the interaction volume was filled with smaller number of balls. This offered support for hypothesis H2, and is understandable because a larger density implies more potential for collisions. These results align with findings obtained by the authors of [71], who employed a similar task albeit with selection rather than retrieval. Secondly, we found that the size of the obstacles did indeed affect performance with users colliding more when the obstacles were larger as opposed to when they were smaller. This confirmed hypothesis H3, indicating that performance of tasks associated with obstacle avoidance, deteriorates when the obstacles are larger. This is rather understandable given that larger obstacles present more potential for collisions simply by virtue of size. Thirdly, we found that the retrieval of deeper targets resulted in poorer performance, supporting hypothesis H5. This isn't surprising seeing as how targets situated at greater depths are blocked by more obstacles. Lastly, contrary to what we expected, we did not find any significant main effect of virtual light intensity on users' performance and hence did not obtain evidence to support hypothesis H4.

We found some interaction effects that further highlight the importance of avatarization. The effect of obstacle density on performance was moderated by the effect of condition. Figure 5 illustrates the effect of density on performance across all the conditions. Compared to both the Iconic and Realistic avatar conditions, the No-Avatar condition experienced a significantly larger detriment to performance for a higher density of obstacles as evinced in the figure. We also found an interaction effect between the size of obstacles and the condition to which participants were assigned. Users in both the avatar conditions performed worse when the obstacles were large, but this difference in performance based on the size of the obstacles was not observed in the No-Avatar condition. Participants in this condition simply performed poorly regardless of the size of the obstacles. Additionally, results on usability gathered from using the PSSUQ revealed that participants perceived the interaction system to be more usable (lower PSSUQ score corresponds to higher usability) when afforded with an avatar. In contrast, not provisioning an avatar was found to be less usable for this task. The findings offer support in favor of avatarization (visualizing the interacting layer) of users' end-effectors in contexts that involve near-field, fine-motor tasks that require high degrees of precision in perception-action coordination.

On examining how users perceived the visibility of their actual hands in relation to the avatar representations and the virtual light intensities applied, we found rather interesting results. With respect to avatarization, the results indicated that users that were not afforded an augmented avatar perceived their actual hands to be significantly more visible than those that embodied an avatar. This suggests that augmenting or overlaying an avatar over a user's real hand, reduces how visible they perceive their actual body to be. This makes sense because the overlaid avatar is likely to obscure one's real hand from being more visible. In terms of the virtual light intensity, users reported that their hands were significantly more visible when the virtual light intensity was low, compared to when the virtual light intensity was high. This suggests that when the virtual light intensity is high, users' real hands are obscured to a larger degree when seen through an optical see-through display. This result aligns directly with prior research showing that as the ratio of virtual to real light intensity increases, the visibility of one's real body tends to decrease [74]. Interestingly, we uncovered an interaction effect showing that the difference in perceived visibility between the two virtual light intensities was less pronounced for users that were provisioned with the Realistic avatar. This is probably because the Realistic avatar obscures the hand to a larger degree than the Iconic avatar whose avatar representation consists of only joints and phalanges. This makes the intensity of virtual light have less of an effect on the perceived visibility of a user's hand for realistic looking avatars. In simpler terms, overlaying a realistic avatar over a user's actual hand, makes it hard for them to be able see their actual hand even under low virtual light intensities. While self-avatarization also reduces actual hand visibility in the presence of Iconic representations, the effect of virtual light intensity appears to be more influential for these kinds of augmented self-avatars.

While we did not see a main effect of virtual light intensity on performance, we uncovered a rather fascinating interaction suggesting that its effect on performance was moderated by the condition to which participants were assigned. In the No-Avatar condition, the number of collisions was significantly higher for a high virtual light intensity than for a low one. In the Iconic avatar condition, however, users performed significantly worse when the virtual light intensity was low. In the Realistic avatar condition, users' performance did not significantly differ based on the virtual light intensity. These effects can be explained based on considering the results of users' conscious reports of their perceived real hand visibility for the two different virtual light intensities applied. When users do not have an augmented avatar (visualization of the interacting layer), increases in virtual light intensity causes detriments to performance because users have to perform the task with less visibility of their real hand and no visibility of the interacting layer whatsoever. This makes it more challenging to perform the task under high virtual light intensities. In contrast, increases in virtual light intensity improves performance in the presence of iconic augmented avatars because the lower visibility of the real hand caused as a consequence, allows users to distalize to the avatar and ignore their actual hand. Reducing the light intensity in such a case, worsens performance by making users' actual hands more visible and potentially distracting. This prevents them from effectively distalizing to the avatar and results in a diminished ability to shift the task-dependent end-effector from their actual hand to the virtual tool (avatar). This idea is underpinned by research showing that in addition to extending the body schema, distalization of the end-effector from the hand to a tool is highly crucial to most effectively exploit the capacities associated with the use of the tool in itself [2]. This was also corroborated by comments made by users in the debriefing interviews wherein users with avatars mentioned that being able to better see their real hands was distracting. When users are afforded with a realistic avatar, the visibility of their actual hands is impacted more by the avatar than just the virtual light intensity applied. Performance in such a case seems to remain unaffected by the virtual light intensity used to illuminate the holograms. These are noteworthy findings that collectively suggest a need to tailor virtual lighting based on whether or not users are afforded with an avatar in AR applications featuring near-field interactions. Apropos of this, it appears that increasing the virtual light intensity is desirable when avatarizing the user, and detrimental to performance when users are not avatarized. Virtual lighting hence seems to be one method to alter the visibility of virtual objects, avatars, and real hands. It is possible that other techniques like contrast highlighting can be employed to erase the real hand from one's view or make the avatar more salient than the real hand.

These results are consistent with the view that the nervous system controls actions via "control laws." These laws characterize how optic (and other) information is used to guide actions such as reaching to a goal, steering locomotion, avoiding obstacles, and tracking moving objects. With control laws, behavior is controlled "on-line" by coupling motor activity to current visual information rather than by generating motor commands from internally constructed mental models of the world and the user's current state [15,81]. In this study, users performed best when the visual information of relevance to such control laws was most salient; information corresponding to the interaction layer (represented by an avatar). While information pertaining to the real hand was less relevant to the this task, higher salience of the real hand improved performance when no avatar was available, and worsened it when users were avatarized. Thus, it appears to be the case that users should be avatarized when interacting with virtual AR artifacts, making the avatar more salient than the real hand. In contrast, when users interact without an avatar, AR developers should ensure that one's real hand is more salient.

## 7 Conclusion and Future Work

In this work, we investigated how avatarization affects users' performance in a near-field interaction task in AR. Users retrieved a target object from a field of obstacles within a box for a number of trials, while avoiding collisions with the obstacles and the box. We employed a between-subjects study design manipulating the presence and anthropomorphic fidelity of augmented avatars. We further examined the effects of obstacle density, obstacle size and virtual light intensity, manipulating these factors within participants. Results indicated that users with tracked avatars performed significantly better than users who weren't provisioned an avatar representation. Furthermore, users perceived the interaction with the system to be more usable when provisioned with an avatar. Additionally, we found that the level of virtual light intensity applied affects the visibility of users' real hands. Overall, we find that avatarizing users' end-effectors, thereby visualizing the interacting layer, improves near-field interactions that require precise fine motor control.

In future work, we wish to investigate how avatarization affects the performance of tasks that involve dynamic components. If similar trends are observed for dynamic scenarios, it would solidify the need for avatarization in AR interactions even more. We also wish to investigate how static self-avatars that are not animated in real time to match users' finger poses affect their abilities to perform interactions. Given that static self avatarization offers solutions to individuals with missing limbs or those that have suffered from strokes, it would be worthwhile to pursue these investigations.

## 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] R. Anderson, J. Toledo, and H. ElAarag. Feasibility study on the utilization of microsoft hololens to increase driving conditions awareness. In *2019 SoutheastCon*, pp. 1–8. IEEE, 2019. 1

[2] M. A. Arbib, J. B. Bonaiuto, S. Jacobs, and S. H. Frey. Tool use and the distalization of the end-effector. *Psychological Research PRPF*, 73(4):441–462, 2009. 3, 9

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

[4] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE virtual reality (VR)*, pp. 3–10. IEEE, 2016. 3

[5] A. Asgary, C. Bonadonna, and C. Frischknecht. Simulation and visualization of volcanic phenomena using microsoft hololens: case of vulcano island (italy). *IEEE transactions on engineering management*, 67(3):545–553, 2019. 1

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

[7] 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. 2

[8] V. Buchmann, T. Nilsen, and M. Billinghurst. Interaction with partially transparent hands and objects. 2005. 3

[9] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn. Fingartips: gesture based direct manipulation in augmented reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 212–221, 2004. 2

[10] 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

[11] 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

[12] S. Coxe, S. G. West, and L. S. Aiken. The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment*, 91(2):121–136, 2009. 6

[13] E. Ebrahimi, L. S. Hartman, A. Robb, C. C. Pagano, and S. V. Babu. Investigating the effects of anthropomorphic fidelity of self-avatars on near field depth perception in immersive virtual environments. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 1–8. IEEE, 2018. 2, 3

[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] B. R. Fajen. *Visual control of locomotion*. Cambridge University Press, 2021. 9

[16] 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

[17] 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. 2, 3

[18] 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

[19] 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. 2, 3, 8

[20] 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

[21] 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

[22] 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

[23] S. Hoermann, E. A. Franz, and H. Regenbrecht. Referred sensations elicited by video-mediated mirroring of hands. *PLoS One*, 7(12):e50942, 2012. 3

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

[25] M. Hoover. *An evaluation of the Microsoft HoloLens for a manufacturing-guided assembly task*. PhD thesis, Iowa State University, 2018. 1

[26] S. Irawati, S. Green, M. Billinghurst, A. Duenser, and H. Ko. " move the couch where?": developing an augmented reality multimodal interface. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 183–186. IEEE, 2006. 2

[27] 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

[28] 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

[29] 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

[30] F. Kaneko, K. Shindo, M. Yoneta, M. Okawada, K. Akaboshi, and M. Liu. A case series clinical trial of a novel approach using augmented reality that inspires self-body cognition in patients with stroke: effects on motor function and resting-state brain functional connectivity. *Frontiers in systems neuroscience*, 13:76, 2019. 3

[31] 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

[32] 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

[33] 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

[34] M. Lee, M. Billinghurst, W. Baek, R. Green, and W. Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Reality*, 17(4):293–305, 2013. 2

[35] 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

[36] C. Lougiakis, A. Katifori, M. Roussou, and I.-P. Ioannidis. Effects of virtual hand representation on interaction and embodiment in hmd-based virtual environments using controllers. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 510–518. IEEE, 2020. 2

[37] 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

[38] 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

[39] E. A. McManus, B. Bodenheimer, S. Streuber, S. De La Rosa, H. H. Bülthoff, and B. J. Mohler. 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*, pp. 37–44, 2011. 2, 3

[40] 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

[41] 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

[42] 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. 3

[43] B. J. Mohler, S. H. Creem-Regehr, W. B. Thompson, and H. H. Bülthoff. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence*, 19(3):230–242, 2010. 2, 3

[44] B. Munsinger, G. White, and J. Quarles. The usability of the microsoft hololens for an augmented reality game to teach elementary school children. In *2019 11th international conference on virtual worlds and games for serious applications (VS-games)*, pp. 1–4. IEEE, 2019. 1

[45] T. Nagel and F. Heidmann. Exploring faceted geo-spatial data with tangible interaction. *GeoViz 2011*, pp. 10–11, 2011. 2
[46] 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. 4
[47] 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
[48] 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. 2, 3
[49] 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
[50] 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
[51] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, G. Lee, and M. Billinghurst. Grasp-shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 73–82. IEEE, 2014. 2
[52] 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
[53] 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
[54] 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
[55] 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
[56] 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
[57] S. R Ness, P. Reimer, N. Krell, G. Odowichuck, W. A. Schloss, and G. Tzanetakis. Sonophenology: a tangible interface for sonification of geo-spatial phenological data at multiple time-scales. Georgia Institute of Technology, 2010. 2
[58] 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
[59] H. Regenbrecht, S. Hoermann, C. Ott, L. Mueller, and E. Franz. Manipulating the experience of reality for rehabilitation applications. *Proceedings of the IEEE*, 102(2):170–184, 2014. 3
[60] 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
[61] 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
[62] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. " these are not my hands!" effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1577–1582, 2017. 4
[63] S. Seinfeld, T. Feuchtner, A. Maselli, and J. Müller. User representations in human-computer interaction. *Human–Computer Interaction*, 36(5-6):400–438, 2021. 1, 3
[64] S. Seinfeld, T. Feuchtner, J. Pinzek, and J. Müller. Impact of information placement and user representations in vr on performance and embodiment. *IEEE transactions on visualization and computer graphics*, pp. 1–13, 2020. 3
[65] D. W. Seo and J. Y. Lee. Direct hand touchable interactions in augmented reality environments for natural and intuitive user experiences. *Expert Systems with Applications*, 40(9):3784–3793, 2013. 2
[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. 2
[67] T. A. Snijders and R. J. Bosker. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. sage, 2011. 6
[68] 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
[69] S. Sørensen and T. Jensen. Development of e-learning applications using hololens and mixed reality. In *European Conference on Games Based Learning*, pp. 649–656. Academic Conferences International Limited, 2019. 1
[70] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE virtual reality (VR)*, pp. 67–76. IEEE, 2016. 2, 3
[71] T. Q. Tran, H. Shin, W. Stuerzlinger, and J. Han. Effects of virtual arm representations on interaction in virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–9, 2017. 3, 5, 8
[72] 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
[73] T. Vuletic, A. Duffy, L. Hay, C. McTeague, G. Campbell, and M. Grealy. Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, 129:74–94, 2019. 2
[74] K. Wang, D. Iwai, and K. Sato. Supporting trembling hand typing using optical see-through mixed reality. *IEEE Access*, 5:10700–10708, 2017. 3, 4, 9
[75] 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
[76] 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
[77] 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
[78] E. Wolf, M. L. Fiedler, N. Döllinger, C. Wienrich, and M. E. Latoschik. Exploring presence, avatar embodiment, and body perception with a holographic augmented reality mirror. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 350–359. IEEE, 2022. 3
[79] H. Xue, P. Sharma, and F. Wild. User satisfaction in augmented reality-based training using microsoft hololens. *Computers*, 8(1):9, 2019. 1
[80] W. Zhang, Q. Zhang, Z. Liu, C. Zhang, W. Wang, and C. Yin. Assembly method for satellite device installation based on hololens. In *IOP conference series: materials science and engineering*, vol. 751, p. 012020. IOP Publishing, 2020. 1
[81] H. Zhao and W. H. Warren. On-line and model-based approaches to the visual control of action. *Vision research*, 110:190–202, 2015. 9
[82] 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