8-2023

All Hands on Deck: Choosing Virtual End Effector Representations to Improve Near Field Object Manipulation Interactions in Extended Reality

Roshan Venkatakrishnan
rvenkat@g.clemson.edu

All Hands on Deck: Choosing Virtual End Effector Representations to Improve Near Field Object Manipulation Interactions in Extended Reality

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Human Centered Computing

---

by
Roshan Venkatakrishnan

August 2023

Accepted by:
Dr. Sabarish Babu, Committee Chair
Dr. Christopher Pagano
Dr. Andrew Robb
Dr. Brygg Ullmer

Abstract

Extended reality, or "XR", is the adopted umbrella term that is heavily gaining traction to collectively describe Virtual reality (VR), Augmented reality (AR), and Mixed reality (MR) technologies. Together, these technologies extend the reality that we experience either by creating a fully immersive experience like in VR or by blending in the virtual and "real" worlds like in AR and MR.

Given the innate potential of these technologies and the impressive rate of technological advancements witnessed in the field, large conglomerates predict XR to transform the future of humanity. Lowering price points and increased accessibility have already led to applying these technologies in several areas, including training, rehabilitation, therapy, collaboration, and education. These XR applications, therefore, exert a tremendous positive impact on our lives.

The sustained success of XR in the workplace largely hinges on its ability to facilitate efficient user interactions. Similar to interacting with objects in the real world, users in XR typically interact with virtual integrants like objects, menus, windows, and information that convolve together to form the overall experience. Most of the interactions in XR tend to occur in the near-field within users' reach envelopes. Many of these interactions involve object manipulation, wherein users manipulate an object in three-dimensional space, picking and placing it around their workspace. A large number of these interactions require precise perception-action coordination and fine motor control.

In XR environments, users are generally provisioned with a visual representation of themselves to guide their interactions. Representations that involve only the distal entity, as opposed to a tracked version of the whole body, are called end-effector representations. The specifics of these representations shape how users perceive XR experiences and further influence their interactions. Given the sparse literature on virtual end-effector representations, this dissertation sets out to investigate how these representations affect the performance of near-field fine motor object manipulation tasks in virtual, augmented, and mixed reality settings.

In our first study, we investigated how different virtual end-effector representations affect the perceptions of dynamic affordances in a near-field object retrieval task in immersive virtual reality. An experiment was conducted in which participants were tasked with retrieving a target object from a virtual box for a number of trials while avoiding collisions with the box's sliding doors. Results of the study indicated that a discordant mapping between the input modality and the end-effector representation produced lower levels of performance. Furthermore, this discordant mapping led to a diminished ability to calibrate interaction performance over trials.

Inspired by previous work showing promise for avatarization in augmented reality, our second study investigates whether provisioning users with a tracked augmented virtual end-effector representation affects near-field object retrieval interactions in an augmented reality setting. In this study, participants were tasked with retrieving an augmented virtual target from a field of obstacles for a number of trials while avoiding collisions with the obstacles. The results of the study suggest that self-avatarization of users' end-effectors improves interaction performance because users have a visual representation of the interacting layer responsible for interactions. Provisioning users with an end-effector representation also affected how visible users perceived their actual hands to be.

The third study in this work investigates the effects of provisioning users with avatars when the interacting components are either physical, virtual, or both. Users performed a peg-transfer task for a number of trials spread out over phases in which the avatar representation and the interaction technique (used to grasp and release objects) were manipulated as within-subjects factors, while the physicality of the pegs was varied between subjects. The results obtained from the study suggest that users were significantly more accurate in the task when the pegs were virtual rather than when they were physical due to the former affording a higher salience of the visual information of relevance associated with the task. From an avatar perspective, provisioning users with co-located self-avatars that are representative of the interacting layer significantly improves their performance. In contrast, provisioning gain-based representations - a technique commonly used to extend reach envelopes and the workspace - tends to degrade performance. In terms of interaction techniques used to manipulate objects, there are accuracy and efficiency trade-offs that designers must consider when developing mixed reality experiences. From a user-experience standpoint, the combination of the end effector representation and the interaction technique used dictates just how usable and taxing mixed reality interactions are perceived to be.

Together, these studies stand to inform the Extended Reality community about how virtual end-effector representations affect near-field interactions and the user experience associated with virtual, augmented, and mixed reality applications. This body of research aids researchers, technologists, and designers by providing them with details that help in appropriately tailoring the right end effector representation to improve upon such near-field interactions in these experiences, thereby collectively establishing knowledge that epitomizes the future of interactions in XR.

Acknowledgments

I am immensely grateful to my advisor, Dr. Sabarish Babu, whose mentorship, scholarly expertise, and unwavering encouragement have left an indelible mark on my intellectual journey. Dr. B, your keen guidance, insightful perspectives, and dedicated support have been the solid foundation on which my academic accomplishments have thrived. Getting to dabble in so many different areas of research, and establishing an eclectic mixture of interdisciplinary collaborations at an international and multi-institutional level has been my favorite part of this journey and this was only possible because of you. I challenge myself to give you a good fight the next time we play badminton!

I am also deeply appreciative of Drs. Andrew Robb, Chris Pagano, and Brygg Ullmer, who generously guided me throughout my doctoral endeavors. As a committee of esteemed faculty members, your assistance has been invaluable at every step, fostering my continuous growth and consistently inspiring me to produce high-quality research. I consider myself fortunate to have collaborated with all of you and eagerly anticipate future collaborations. Dr. Pagano, thank you for introducing me to the ecological approach to perception and breaking it down as how one would to a ten-year-old. It really has changed the way I look at things, especially beyond the scope of research, so much so that I promise to erase the word 'cue' from my vocabulary. Dr. Robb, your affable nature and amiable demeanor makes it so easy to seek your technical guidance, something that I will strive to achieve wherever I am in the future. Your amazing baking skills make me think your post-retirement plan should involve a patisserie in there somewhere. Dr. Ullmer, your creativity, originality, and expertise has greatly helped broaden the scope of my research providing futuristic implications on how society will one day use technology based on my findings. As your neighbor, I'm glad I frequently run into you and promise to never be noisy or ever call the cops on you (not that I ever have to) :-)

I would like to extend warm acknowledgment to four other Venkatakrishnans who have played a significant role in my journey. First and foremost is to my partner in crime, my twin brother, Rohith, who has been my constant companion throughout life. I'm grateful to have navigated this scholarly journey together, collaborating synergistically on numerous research projects while continuously motivating, encouraging, criticizing, and entertaining each other with all the goofing around. Rohith, If the two of us of all the people on this planet can earn a doctoral degree, then I'm genuinely worried for the world and the future of research. Anyways, grab your cleats, and let's go play 5 v 5 this evening and I'm not defending! The next Venkatakrishnan deserving of a shout-out is my venerable older brother, Roopak, who has been a guiding light throughout my life. With Roopak by my side, I can confidently face even the most challenging problems, knowing that a solution can always be found. For all the assistance he has provided, day or night, I will forever remain indebted to him. His support, understanding, and encouragement have been invaluable, and I often say that one couldn't ask for a better older brother. Roopak, when your company makes it big, please hire me - I will definitely continue to annoy you. The other two senior Venkatakrishnans who deserve credit for my accomplishments are my parents. You guys have believed in me even when I showed no signs of responsibility in life. Your unconditional love and faith in my abilities have been a driving force behind my achievements. Thank you for always being there to cheer me on. As kids, Rohith and I were always about sports, goofing around, and doing things that not many parents would support. Mom and Dad, keep up this trend and don't change because "we are primordially circuited to..."

I am extremely thankful to a very close friend, Michael Stamatakis, for having planted the idea of me pursuing a doctoral degree. I still remember how we became friends in that one graduate class where we simply hit it off, our age differences simply being non-existent. It took a couple of classes, group projects, presentations, IT work together, and time to upgrade that friendship to a landlord-tenant relationship. I still love that I have to remind you about rent. Michael, never once did I think I'd be thanking my landlord, but you sure are one hell of a landlord, friend!

I would like to extend a special thanks to Ryan Canales and Balagopal Raveendranath for their invaluable assistance in my research projects and friendship beyond that. Ryan, keep the "Eugene moments" coming because they are "low-hanging fruits" right Bala? Additionally, I am deeply grateful to my friends Ayush Bhargava, Alex Adkins, and John Porter III. Ayush, only you can be a ninja and Dr. Super at the same time; Alex, let's hike table rock again? I promise I'll try doing it with water this time; John, let's settle this once and for all- you're the king and I'm big chief. Here's also thanking Dane Acena and Catherine Barwulor for providing great entertainment throughout this journey. All of you have collectively created a light-hearted atmosphere and made this shared journey enjoyable. Our friendship has been a source of great camaraderie, laughter, banter, and unwavering support. The bond we have forged is undoubtedly one that I will endure for the rest of my life and I'm glad we went through this together.

A special mention goes to Dr. Mark Smotherman, Matt Dabney, Chuck Cook, Sasha Kuksenok, Christy Babb, and Kaley Goodwin from the School of Computing. Thank you all for your support as staff. Here's to remembering all the laughter, food, pranks, and jokes we've enjoyed together on a daily basis, and working with you all and everybody else from the School of Computing has been nothing but a pleasure.

Last but not least, I want to express my gratitude to 'The Room'. This magical space has witnessed it all - from post-surgery rehabilitation to intense research-related discussions, from moments of music and solace to playful interactions with all the beloved dogs we've watched over, from cherished hangouts, laughter, and conversations with friends to moments of peaceful meditation. This room has been my sanctuary, where ideas come to life and where the spark of inspiration finds its home.

In conclusion, this dissertation would not have been possible without the support, encouragement, and contributions of the incredible individuals mentioned above along with many other friends who have supported me in ways beyond words. Thank you all. And remember, while I may still show no signs of responsibility in life, I assure you that I take my people seriously! **ALL HANDS ON DECK!**

Figure 1: The Room

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Extended reality, or as most people prefer to call it by its acronym, XR, is an emerging all-encompassing term that refers to a family of multi-modal immersive technologies in existence today. These include the popular virtual reality (VR), the rapidly growing augmented reality (AR), and mixed reality (MR). To experience these immersive technologies, at least from a contemporary standpoint, users commonly wear what is called a head-mounted display (HMD), which serves as the primary device for these technologies. In all three of these immersive technologies, the virtual content and information provided can span multiple sensory modalities, including but not limited to visual (most common), auditory, haptic, olfactory, and somatosensory channels. During this last decade, these technological devices have seen tremendous advancements and improvements with respect to their form factors, tracking fidelity, display characteristics, and graphical capabilities. With continued competition among manufacturers of these devices to be able to provide readily available and affordable consumer-grade devices, the total market size for XR across the world was estimated to be roughly 28 billion US dollars in 2021. This rather conservative estimate is projected to hit around 300 billion US dollars by the time 2024 comes around [9].

Even with all this growth, there continues to be a fair amount of confusion, even among experienced individuals and technologists, as to what exactly the differences among VR, AR, and MR are. It is hence important to clearly distinguish the terms before proceeding any further. Virtual reality, on the one hand, allows users to be immersed in artificially generated virtual worlds. This ability to experience simulated virtual environments with highly realistic graphical renderings opens up a multitude of possibilities for applications to make use of. Many big tech companies are already pushing for the ubiquity of VR devices and their adoption through what is called the Metaverse, a single universal immersive world as the next big iteration of the internet (especially after the pandemic). Currently, VR already finds its use in areas including training [78], therapy [223], education [39], collaboration [166], perception research [123, 124], sports [127], driving and rowing simulators [214–217], and extreme activities [107] to name a few. Augmented reality, on the other hand, provides an immersive (though less immersive than VR) experience by overlaying virtual content in the real world. Its name derives from the fact that the virtual content in AR is 'augmented' to the real world we continue to experience. The overlaid virtual information can be either constructive (adding to parts of the real world) or even destructive (obscuring parts of the real world). This capability of being able to augment (superimpose) information and digital content to reality creates a number of applications that warrant its use. Many domains including medical training [13], shopping and retail [91], education [113], and manufacturing [83] are already starting to benefit from using AR applications. Mixed reality furthers this concept by bundling in an interactive physical component to the experience. In addition to being able to see overlaid virtual artifacts in AR, it is possible for physical components to be able to interact with virtual components. Essentially MR breaks down the barriers between physical and virtual realities, blending in the physical and digital information provided while interacting in the medium. The concept of the reality-virtuality continuum (also referred to as the Mixed-reality continuum by some researchers) was first introduced Paul Ingram; This continuum encompasses all possible compositions and variations of real and virtual elements, clearly delineating reality (completely real) from virtuality (completely virtual) [146]. The region in the middle (MR) of this scale includes augmented reality - where virtual elements augment or add to the real world, and augmented virtuality - where real elements augment the virtual. For all practical purposes, the terms MR and AR continue to be used interchangeably by researchers, academics, and technologists.

The continuing success of these applications largely depends on the advancements made with respect to user interaction or how the end-user or consumer of the technology interacts with the system(s) and its related aspects. This is not specific to any one of the XR mediums (VR, AR and MR) but all of them together. Just as how interaction is central to our experience in the real world surrounding us, they strongly influence our perceptions of the experience in these immersive mediums, thus making their implementational mechanics highly germane. Users in XR typically interact with digital or virtual integrants, including objects, information, menus, humans, interfaces, etc., that, together, convolve to form the whole experience. For example, one can interact with a virtual human in VR [221], see the weather information overlaid onto their field of view when wearing AR goggles, or even collaborate on a project with other users in MR [152]. Moreover, recent developments in XR allows users to be able to interact with virtual components using physical ones as well. These include digital pens [85], 3D printed tangible entities [51, 68], and one's actual hand(s). One way to categorize interaction in such XR settings is based on the region or space at which the actual interaction takes place in relation to the position of the user [48]. The vista space or far-field tends to include interactions that take place beyond thirty meters from the user. The action space also called the medium field, involves the interaction in regions that occur between six and thirty meters from the user. Lastly, the personal space or near-field tends to, as the name suggests, include interactions that occur within the maximum arm-reach of the user or just slightly beyond their reach envelope.

Most of all the interactions in these extended reality mediums tend to occur in the latter regions wherein users interact with virtual artifacts in their peri-personal space. Picking up virtual items, moving them around, moving menus from one side to the other, and dragging and dropping virtual windows are among many other forms of such interactions. Supporting near-field interactions faces many a number of hindrances and depends on multiple factors that together affect how well the interactions take place. The tracking capabilities of the HMD, the latency associated with tracking users' hands, and the lack of haptic feedback are some examples of problems that continue to pose challenging questions to the goal of being able to achieve flawless interactions. Regardless of the status quo from a technological standpoint, near-field interactions in even the real world can be quite complex. Threading a needle requires an individual to correctly feed the thread through the needle's eye; connecting electrical circuits involves the arrangement of wires, switches and other components in a carefully systematic fashion; operating on a patient during surgery hinges a lot on being able to use surgical tools like scalpels, forceps, and retractors effectively. Such types of personal space experiences in XR are just as complicated, if not more, given the challenges in facilitating these kinds of interactions from a technological standpoint. It hence follows that precise and accurate fine motor perception-action coordination is required to be able to effectively and efficiently interact with virtual artifacts in the near-field. Moreover, with applications in these mediums already seeing their integration into industrial, educational, training, and medical settings especially, it is imperative to be able to design these systems to support effective user interaction in the near-field.

Being able to manipulate objects is the most effective way to interact in the near field. Object manipulation revolves around one's ability to, as the name suggests, manipulate objects in space using their own hand(s). This involves moving, positioning, and placing objects with one's hands and limbs (end-effectors) without the help of other influences. We as humans manipulate objects in the near-field on a daily basis and it forms a central part of our functioning lives. Quotidian examples include writing using a pencil, pouring water out of a jug, cutting with a pair of scissors, wielding a rod, picking up a smartphone, and using a pair of keys, to name a few. The ability to perform these tasks relies on the coordination of many skills, including but not limited to hand-eye coordination, fine motor control, and bilateral integration. Object retrieval can be construed as a type of manipulation requiring the user to retrieve an object using their hand or end-effector(s). Being able to retrieve objects allows one to physically inspect (and, as a result, interact with) objects in one's vicinity or reach. Like in the real world, object manipulation can be facilitated in XR settings wherein a user interacts with (manipulates) a virtual entity. Most research with respect to object manipulation in XR is looked at through the lens of 'pick and place' tasks, where users both retrieve and place an object from a source to a destination [7, 130, 196]. Of all the near-field object manipulation techniques, grasping virtual objects is one of, if not the most fundamental techniques, regardless of whether it is VR, AR or MR. It is the motor process that allows one to hold (onto) something and can be considered as an egocentric manipulation technique that allows for a virtual object to be manipulated. The mechanics of near-field interaction paradigms vary greatly depending on the medium in which interaction is taking place. In VR, interactions are usually facilitated by the provision of handheld physical controllers (Oculus touch, HTC Vive, Valve knuckles, etc.) that accompany the HMD or using other tracked physical devices like motion capture suits and high-fidelity gloves. Additionally, camera vision-based gesture tracking controllers (Leap motion, Microsoft Kinect, etc.) can also be configured to work with the VR system. In AR and MR, however, users typically rely on hand-tracked gestures and actions. This isn't to say that it is not possible to interact using physical controllers in AR or MR, but that it is less often done. Thus, designing efficient near-field interactions will hence vary depending on the medium in which the user is immersed or rather interacting in.

The term 'Representations' holds significant meaning across various disciplines, including art, math, politics, science, and computing to name a few [115]. In art, representations refer to the visual or conceptual portrayal of ideas, emotions, or experiences through various mediums such as paintings, sculptures, or installations. They serve as powerful tools for communication and interpretation, allowing artists to convey their perspectives and invite viewers to engage in dialogue. In math, representations involve the translation of abstract concepts into tangible forms, such as numbers, algebraic equations, or geometric figures, enabling mathematicians to explore and analyze complex mathematical structures [70]. In politics, representations encompass the ways in which individuals or groups are depicted and advocated for, fostering inclusivity, equity, and participation within democratic systems [170]. In science, representations involve models, diagrams, or simulations that aid in understanding and explaining natural phenomena, providing visual or conceptual frameworks for scientific inquiry. Lastly, in computing, representations refer to the encoding and manipulation of information in various formats, such as binary code or symbolic notations, enabling the creation and operation of digital systems that underpin modern technology and communication [154]. Across these disciplines, representations serve as vital tools for expression, analysis, comprehension, visualization, and innovation, shaping our understanding and interaction with the world around us.

In XR, representations take on a new dimension by immersing users in simulated environments. They provide a bridge between the real and virtual worlds, allowing individuals to interact with and experience digital content in an immersive and interactive manner. In such virtual experiences, users are often provided a visual representation of themselves, and these are called user representations [193]. Such representations allow users to be able to interact with the virtual artifacts in the environment. The most common example from a non-immersive standpoint is the mouse on a desktop computer. Considering it as a simulacrum of sorts, the mouse observed on one's desktop monitor visually depicts the spatial correspondence of the user in the environment, in this case, the screen and what it currently displays. The mouse is what facilitates interaction by allowing users to click on icons, draw on the screen, drag objects around, scroll through pages, and open/close windows. In XR, however, this depiction of the user is no longer relegated to a two-dimensional, third-person perspective (3PP) of one's self on a screen. Instead, users in these immersive mediums are provided with a visual representation that depicts them in relation to the three-dimensional world they are currently immersed in, typically from a first person perspective (1PP). Such representations can also be achieved from a 3PP, but more emphasis is placed on the former. These user representations are commonly referred to as self-avatars, and they have been shown to affect how users perceive and interact in XR experiences [17, 55].

From a visual standpoint, users can be provisioned with full-body avatars in both VR and AR/MR. In VR, this may be achieved using optical tracking systems with motion capture suits or by attaching trackers to different parts of the body. Using optical tracking to provide full-body self-avatars requires highly expensive infrared (IR) cameras to track hundreds of markers affixed to the user (on their clothes or motion capture suit). Moreover, the lack of physical space, requirements of complex scripts to track users in real-time, and long set-up times make the provisioning of full-body avatars using such techniques highly challenging. Enter contemporary trackers like the HTC Vive trackers, for instance, which rely on advanced sensors like gyroscopes, accelerometers, lighthouse emitters, and IMU sensors, making it much easier to avatarize users' full bodies using inverse kinematics. However, using trackers is not as accurate as optical systems simply because of the reduced number of data points for tracking. Provisioning full-body avatars in AR/MR depends on how the system is realized. In optical see-through (OST) immersive AR, full-body avatars can be achieved based on the positional tracking of the HMD or by using a holographic mirror. Video see-through (VST) immersive AR relies on additional trackers, controllers, and cameras. As such, provisioning full-body avatars in XR requires higher tracking capabilities and can suffer from latency and other tracking errors leading to detriments in interaction performance. Besides, such problems also tend to negatively affect user perceptions and the overall experience in general [143].

Given the added hardware requirements and computational complexity associated with tracking and provisioning full-body avatars, it is common to only track users' end-effectors, or more simply stated, the entities that are directly involved in near-field interactions. The term end-effector representation can hence be construed as a subset of user representations wherein the user is represented simply by their end-effector rather than their full body. This is very common in VR wherein one's end-effectors are mostly often depicted as virtual replicas of tracked handheld controller models, task-dependent tools (e.g., hammer, forceps, scalpel, and drill) depending on the scenario, or human-like hands depending on which exact entity (visualized) is directly responsible for the interactions that occur in the peri-personal space. These representations have been shown to affect user experiences (enjoyment, presence, perceived realism) in VR [1, 121]. In attempting to maximize users' sense of embodiment, many VR experiences are resorting to visualizing users' end-effectors as human-like hands as opposed to virtual replicas of the handheld controllers themselves. For example, experiences in the Oculus Quest 2 HMD tend to visualize users' end-effectors as hands (either tracked based on camera vision or when using controllers). Little is known about how such virtual representations affect perception-action coordination with respect to interaction. The paradigm of AR, however, was conceptualized, allowing the real world to remain visible, logically implying that one's real physical body would also be visible. The fact that visual information about one's physical (actual) body is still salient in AR/MR would make end-effector representations seem unnecessary and inconsequential to interaction. This being said, avatarization in AR continues to be explored given the potential it offers for embodiment, and interaction [67, 186]. Along these lines, self-avatars (end-effector representations) have been proven to be useful for medical prosthesis [200], recovery and rehabilitation [77, 98], and physical performance [160]. Many such examples lend support for self-avatarization in AR. It remains to be seen if and how provisioning such end-effector representations affect interactions with virtual entities in AR. Furthermore, with the majority of modern AR applications involving some form of near-field interaction with virtual artifacts using hand-tracked gestures, it is more than imperative to understand how augmented self-avatarization impacts the effectiveness of near-field interactions.

Practically speaking, while virtual hand representations are also end-effector representations, the latter seems to be a more appropriate term given that users in VR are not always represented as human-like hands. Regardless of the niceties surrounding the nomenclature, it is the distal virtual entity (tool) that dictates the potential for interactions, and it is this entity whose virtual representation is the subject of discussion. Literature on how tool use affects one's body schema is abundant, with research continuously showing that tools have the ability to extend the perception-action capabilities of an organism [137]. Once picked up, the tool shifts the boundary between the actor and the environment, thus affecting the perception of reach or distance [178, 232]. This fact also holds true for virtual tools in immersive mediums [50]. Users incorporate these end-effector representations effectively into their body schema over time, thus affecting their perception-action in the medium as well. Given the sparse literature on the effects of end-effector representations in near-field object manipulation interactions, this dissertation sets out to investigate how virtual end-effector representations affect the performance associated with near-field fine motor object retrieval tasks in virtual, augmented, and mixed reality settings. Seeing as how XR is now increasingly being used in industrial, operational, medical, and other training-related settings, wherein objects within one's reach need to be carefully and precisely retrieved, knowledge about user behavior and perceptions as a result of these representations is critically dissected and further discussed from an interaction and user experience standpoint.

Chapter 2

Related Work

2.1 User Representations in Virtual Reality: Interaction and Embodiment

In immersive virtual environments (IVEs), the user is often provided with virtual representations that are commonly referred to as user representations [193, 194]. They are also commonly referred to as self-avatars. These representations help users perform actions and affect how effectively, accurately, and efficiently users can perform them [54, 142, 148, 205]. Furthermore, they play a key role in shaping users' sense of embodiment, a phenomenon that comprises the sense of self-location, agency, and ownership [102, 129], which are important characteristics associated with interactive IVEs. De Vignemont [52] taxonomizes the sense of embodiment into three dimensions (Spatial, Motor, and Affective) that directly relate to the senses of self-location, agency and ownership respectively. Agency corresponds to motor activity control over a body part that obeys one's will and sensation of movement [22], while ownership pertains to one's self attribution of a body [66, 210]. User representations can differ in terms of their visual appearance, the input modality used to support their actions, and the mapping of their control mechanisms [194].

On the subject of the different input modalities used to support interactions in VR, several systems have been implemented and further investigated in terms of how they are perceived by users and how much they affect task performance. Interactions can be facilitated through direct VR handheld controller inputs, camera vision based tracked hand gestures, using magnetic or reflective markers with a motion capture system, or even through supplemental hardware that have been configured with the system. Examples of these supplemental input modalities include the Leap motion controller (though this is also a camera vision based tracking solution), optitracked gloves [138], Noitom's Hi5 VR gloves with inertial measurement unit (IMU) sensors, etc., that use additional hardware to facilitate hand tracked gestures for interacting with the IVE. These input systems typically differ in the means through which they support interactions, their tracking and latency characteristics, and more broadly, their accuracy and fidelity. Ultimately, these factors affect users' ability to perform tasks in the environment and invoke different levels of perceived agency, ownership and realism associated with their representations in an immersive experience. Research on this front that compared the use of the HTC Vive controllers to the leap motion controllers in performing motor tasks (i.e, selection, placement and rotation) found that the former input method resulted in faster performance and lower perceived difficulty [35]. The visual and kinematic properties of a virtual hand representation and its corresponding motor synchronicity can also affect task performance [38, 159]. Other studies have shown that virtual hands do not significantly impact performance for tasks involving tool based object manipulation [128, 183].

Interactions through direct hand gestures (e.g., using tracked data gloves, visual gestures recognition, etc.) tend to be associated with higher levels of body ownership and perceived realism than interactions performed using physical hand-held controllers [121]. Users in the aforementioned study preferred interacting using tracked gloves over the use of controllers despite the latter resulting in better task performance. Other works have obtained similar results wherein gesture based interactions supported using tracked gloves or hand tracking were preferred over controllers despite resulting in diminished performance [141, 190]. On a similar vein, it was found that using hand tracked gloves to support interactions in IVEs was associated with higher levels of ownership, realism, enjoyment and presence than when using controllers [1]. This study however did not investigate task performance. Research has also shown that users tend to pay more attention to tasks when performing them based on direct hand inputs over the use of controllers [5]. A recent study comparing interactions supported using controllers against hand tracking found no significant effects of providing users with tracked hand gestures on the perceived usability and satisfaction associated with the representations [101]. Work that investigates near field writing on-air suggests that users write slower when using a pointing gesture than when using a controller to write [218]. Overall, research suggests that input modalities that support hand tracked gestures in IVEs seem to be associated with higher levels of embodiment, naturalness and perceived realism over physical hand held controllers despite the fact that using such controllers tend to produce better performance [147, 198].

With respect to the visual appearance, user representations can range from the provisioning of realistic high fidelity self avatars to the user being represented through virtual models of the controllers they physically hold in their hands. To support realistic avatars, body tracking technologies are used due to their ability to allow users control movements of the virtual avatar based on their own movements. Such technologies promote a high sense of agency and embodiment seeing as how users perceive and control their virtual body as if it were their own [8, 102, 105]. A number of studies have demonstrated that interacting with IVEs through virtual avatars produce more favorable outcomes. Along these lines, work conducted by Peck et al. has shown that appropriate user representations can help reduce racial bias [164]. Research further suggests that users experience lower cognitive loads while performing a spatial rotation task in the presence of self avatars than without [205]. Furthermore, the provision of a fully tracked avatar was found to produce more accurate egocentric depth judgements when compared to judgements provided in the absence of such an avatar [148]. Depth estimations were also shown to improve when users were provided with high-fidelity self avatars, compared to low-fidelity avatars or seeing the virtual end-effectors alone [54]. The realism, size, and shape associated with one's own self avatar has also been shown to affect the perception of the size of virtual objects [125, 155]. Recent work has shown that higher visibility and anthropomorphism of self-avatars results in users adopting more realistic behaviors (i.e., not penetrating a virtual wall) in IVEs [157].

2.2 End-effector Representations in Virtual Reality: Near Field Interaction and Embodiment

End-effector representations can be considered a subset of user representations which involve the representations of the hands or the tools with which actions are performed in the virtual world. These end-effectors refer to the virtual tool or the part of the representation that comes into contact with the object being manipulated [6]. Typical ways in which contemporary VR experiences represent end-effectors include rendering them as virtual models of the handheld controllers themselves (virtual replica), or by rendering them as virtual hands that may vary in realism, size, and fidelity, or by depicting them as a tool. Similar to user representations, the way in which end-effectors are represented affect task performance, embodiment and other characteristics associated with interactions in IVEs. A few studies have investigated the effects of the visual appearance of end-effectors on task performance and embodiment while managing to keep the input modality consistent. Along these lines, in a study leveraging the HTC Vive controller as an input modality, it was found that visually representing the end-effectors as hands leads to increased levels of body ownership than when rendering them as controllers or spheres [130]. The authors of the aforementioned study found that the controller and hand representations produced better performance in a pick and place task than the sphere representation. Their results additionally revealed that visually representing the end-effector as a controller was better suited for a positioning task as compared to representing the end-effector as both a hand, and a sphere. In another study that used the leap motion controller to provision three gesture tracked virtual hand representations (i.e., abstract, iconic and realistic), the authors found that the sense of agency was higher for less realistic virtual hands owing to the lower degree of mismatch between the users' gestures and the animations on the virtual hands [8]. In contrast, a higher sense of ownership was associated with the realistic hand representation. In terms of task performance, it was found that simplified end-effector representations produced faster and more accurate interactions for a pick and place task. In a study comparing three representations of virtual arms (i.e., 'hand-only', 'hand+forearm' and 'whole arm") on the performance of a selection task, it was found 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 [209]. One other study investigating the effects of end-effector representation on body ownership, immersion, perceived difficulty and performance found no differences between representing hand held controllers as virtual hands, virtual controllers or partially rendered virtual bodies [135]. Overall, studies that have investigated how end-effector representations affect interaction and embodiment have focused on tasks that involve selection, placement, collision avoidance, etc., in the near field where the affordances are relatively static. The effects of end-effector representations on interactions that involve tasks with near-field dynamic affordances seems to be an avenue of research that remains relatively unexplored.

2.3 User Representations and the Perceptions of Affordances in Virtual Reality

Affordances refers to what the environment offers the individual. They represent the relationship between the properties of the organism and characteristics of the environment and can be described in terms of the organism's own intrinsic units [23]. Stated simply, they refer to the action capabilities of the perceiver in the surrounding environment. Our lives expose us to a multitude of affordances, quotidian examples of which include passing through openings such as doors, hallways, etc., grasping objects, sitting on chairs, and stepping on stairs to name a few. Scaled to the organism (actor), affordances are determined by the morphology and physical capabilities of the actor. The possibilities for action that arise from the relationship between the environment and the geometric scale of the perceiver is described as body-scaling in affordance literature [89]. For example, the shoulder width of an individual determines their ability to pass through an aperture [226, 227]. Action-scaled affordances furthers this concept by including both kinematic and kinetic abilities of the body to act in dynamically changing environments like catching a moving ball [59, 187]. It hence follows that manipulations of an individual's body schema or their action capabilities affect the way they perceive the environment [206, 227]. This has also been demonstrated in contexts of virtual environments wherein manipulations of the user's self-avatar or their action capability have been shown to affect perceptions of depth, size, weight, and passability [54, 155].

One way to classify affordances is based on the dynamicity of the affordance or environment thereof. Research on static affordances wherein the environment remains relatively static has been widely studied in the context of virtual environments. For example, it has been shown that foot size of a self-avatar in an IVE affects judgments on whether gaps can be perceived as crossable [96]. Similarly, the appearance of users' self-avatars (i.e, overweight/underweight) was shown to affect passability judgments for apertures created between two poles, demonstrating that the size of the user representation affects how an affordance is perceived [169]. In these studies, the affordances of the gaps and apertures are static in that they exist in an environment with unaltered spatio-temporal dynamics. However, interactions can also involve affordances that exhibit dynamic properties. Examples of such include crossing gaps in traffic and crowds, kicking a soccer ball, picking up luggage from a carousel, etc., wherein elements of the environment move. The spatio-temporal dynamics of such environments require the actor to synchronize their own movements to the movement of the object(s) in the environment, making these affordances more complex in nature. Some studies have explored dynamic affordances in the real world. On this front, investigations have explored dynamic affordances in contexts of gap crossing, walking through oscillating or closing doors [33, 45, 60, 132, 134], and crossing streets with oncoming traffic [42, 158, 174]. In IVEs, such research has shown that the locomotion method (action capability) affects the perceived opportunities for action in the environment [71]. Participants in this study were tasked with choosing from among a series of opportunities to pass through a gate that cycled open and close and then board a moving train. The mode of locomotion was varied between walking and joystick control, and was tested for both HMD and CAVE displays, with the results indicating that both manipulations affected performance. The affordance in the aforementioned study was dynamic such that participants had to make judgments of passability through an aperture that constantly varied in width. With an intention to improve traffic crossing behavior and safety in pedestrians and bicyclists, IVEs are frequently used to study such affordances [20, 42, 174, 197]. Although dynamic affordances have been examined in IVEs, the relationship between user representations and their concomitant effects on such affordances remains relatively unexplored. Given that user representations affects the perception of static affordances in IVEs [18, 19, 55, 117, 148], it stands to reason that such visual representations of the user could in fact affect how users perceive affordances in more dynamic contexts and scenarios in immersive virtual environments.

2.4 Body Schema and Embodied Interactions

The concept of a body schema has been around for more than a century now [75, 76]. The idea purported by the term schema is that it is a representation of one's body dictating the potential for performing actions. Considering that organisms like ourselves grow, physically, with time (age), it is reasonable to expect that one's body schema would change as a result. While early theorists argued that changes to the body and its concomitant action capabilities are compared to an internalized body schema based on memory of this representation, more recent work suggests, and also tends to contest that it is malleable and is continuously perceived when moving and equipped with objects [162]. Support for this argument stems from the fact vibrations to the muscle tendons can induce the illusion of limbs in impossible positions, from an anatomic standpoint at least. What this rather suggests is that proprioception, or the perception of limbs in space, happens on-line rather than based on a stored representation of the schema accessed through memory. The body schema adapts to one's body, regardless of whether the changes are temporary, or even permanent. Two important terms that tend to be used interchangeably in the literature that directly tie to this concept are adaptation and calibration [149]. The former term tends to be phrased around the idea of adjusting to perturbations and manipulations of embodied units [21, 46]. The latter meanwhile, has in the past been framed around the process of actors or organisms, adjusting to changes to the action units. The term attunement also tends to find its place in the literature, referring to the process by which organisms become sensitive to information (variables) that potentiates the perception of pertinent affordances [30, 204]. As such, there isn't a clear cut demarcation between the terms among researchers in the community, allowing one to come away with the conclusion that both adaptation and calibration relate to the processes associated with perception-action learning.

The properties of the organism that affects the perception of affordances, and also shows the reciprocity of the relationship between the organism and the environment, is what literature refers to as effectivities [211]. What effectivities of an organism mean for acting is commensurate to what the environment offers for perceiving an action capability. In other words, the term effectivity boils down to a means of acting that allows an individual to realize a specific affordance [79]. Consider a branch of a large tree. What one person perceives as a branch that affords hanging, by virtue of their athletic arms and body, is perceived by another individual who isn't as strong and athletic, as choppable using a chain saw. Perception occurs in the two organisms in the same manner- attunement and calibration to ambient energy arrays to directly perceive the affordance. The two individuals perceive the same world based on their respective effectivities, in this case, their strength and athleticism or lack thereof.

Using a tool extends the existing perception-action capabilities of an organism, and this is true, especially for humans. Holding a hammer amplifies the potential for striking while using a pair of binoculars amplifies one's visual ability. Prior to use, a tool is simply a detached object that exists in the environment or rather an extension of the environment therein. Once picked up, it can be treated as a functional extension of the user, crucially playing a role in extending the effectivites of the individual that is using it. Using a tool hence shifts the boundary between the organism and the environment [137]. Human beings have quite the ability to assimilate tools (being used) to factor in and extend reach into their body schema. This in turn ends up altering how they perceive depth or distance [178, 232]. Just like in the real world, virtual tools (user representations/ avatars and end-effector representations) constitute tool use in the virtual world, and directly affect perception in the same manner in which it does so in the actual world [50]. A Multitude of studies more than demonstrate this phenomenon in VR wherein the anthropomorphic and anthropometric properties of virtual representations being used in the virtual world affect distance perception, passability perception, and size perception to name a few [17, 55, 155]. Something similar can be said of augmented reality where a relatively fewer number of studies demonstrate the same [62]. What sets itself as a mutually agreed upon distinction is that users attune and calibrate to manipulations and perturbations to the body schema over time, thus affecting their perceptions and interactions in these mediums as well.

2.5 Interaction in Augmented and Mixed 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 [11]. 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 [230]. Researchers have continued to extensively explore such modalities to support intuitive and immersive interactions with virtual entities in the near field [7, 237]. The findings from a majority of these research efforts suggest that users tend to be more efficient in performing near field interactions though direct manipulation, expressing preference for modalities that are supported through natural means like freehand gestures [32, 72, 173]. However, there have been some works that have found preferences for non-natural interaction systems over natural ones [177].

Hand-based AR interactions often require users to perform gestures like tapping a marker [41, 150], making a fist or opening the palm [175, 181], pinching or pushing with the finger tip [150, 172], or manipulating an object through space [56, 180, 191]. While some systems limit interactions to one hand [144, 181], 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 [131, 161] or by using depth cameras, video cameras or infrared sensors [110, 195, 212, 222]. 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, perceived 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 [61]. Furthermore, far-field interactions may manipulate geometries that extend beyond the user's arm reach [80, 167]. 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 [106]. For these reasons, multi-modal interaction modalities that combine technologies like speech, gesture recognition, etc. have been introduced and researched [90, 225]. In comparing free-hand gesture-based interactions and multi-modal gesture-speech interactions, the strengths and weaknesses of these systems have been discussed. Free-hand gestures tend to support better spatial input, while speech commands offer better system control [88, 114]. 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 [171]. Other research efforts comparing these interaction techniques have found that speech tends to outperform gestures in terms of accuracy but that the simplicity of gestures more than compensates for this loss in accuracy with speed [40]. 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 [97].

The manipulation of virtual objects using hand gestures in AR can involve two methods namely metaphoric and isomorphic hand interactions [136]. 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 that base input actions 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 [64]. Other researchers have also found no differences between these interaction methods in terms of both task performance as well as users' subjective perceptions [196].

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 [12]. 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 [231]. 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 [145]. This being said, near-field interactions in AR that involve the 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 gesture to control virtual objects in a short amount of time [179].

2.6 Avatarization in Augmented and Mixed Reality: Embodiment and Interaction

Avatarizing users in augmented reality is gaining popularity given the potential for users to perceive embodiment like in VR. This opens the door to a numerous number of applications in field such as medical practices [112], education [94], remote collaboration [152], video games [185], 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, the user perspective (1PP and 3PP) leveraged , etc. Optical see-through HMDs overlay augmented avatar content on the real world while video see-through 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 [233]. Similar to the reality-virtuality continuum [146], 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 [67]. 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 [200], and rehabilitation and recovery [77, 81, 98, 182]. Research on manipulating the appearance of the users' self avatar in AR shows that when embodying a more muscular avatar, users physical performance improves [160]. Recent work that has investigated the effects of hand representations in AR has shown the feasibility of using AR avatar arms that are expandable to interact with far-field real word objects connected to the system [62]. In this study, the 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 have been made to study how user or end-effector 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 interactions in AR. Thus, while this topic is widely studied in VR, it remains a young research field in AR.

2.7 Impact of Real Body Visibility on Avatarization in Augmented and Mixed Reality

Unlike immersive virtual reality, using an augmented or mixed reality 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 affect 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 [224]. The results of their work suggest that real body visibility need not necessarily impede embodiment experiences from a first-person perspective as long as the virtual hands are more visible than the users' real hands [67]. In terms of the transparency of the real hand (real hand visibility), users prefer performing interactions 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) [31]. Users in this study found it odd and absurd to interact when their hands were less visible, commenting that it was alien-like, weird, etc. Furthermore, it was observed that the perceived level of transparency varied depending on the color of the background that the hand was being viewed against. However, it must be noted that users in this study were not provided with a self-avatar or a visual representation of themselves. Overall, the impact of real body visibility on interaction performance in augmented and mixed reality settings is an area that warrants more concrete investigations.

2.8 Object Manipulation in Virtual, Augmented, and Mixed Reality

Interaction with virtual objects and manipulating them through space is a topic that has been extensively explored in the near field [7, 73, 92, 190]. Grasping is a common interaction that people perform routinely and its realistic simulation in VR requires dedicated hardware and algorithms [24]. It can be treated as an egocentric manipulation technique when simulated using virtual hands which in turn allows users to hold, move, drag and constrain virtual objects upon contact with the hands [176]. While visualizing grasping, depicting the grasping entity (hands) and allowing it to penetrate the object can result in better performance, embodiment, and enjoyment than when disappearing it or snapping the object to it [1, 37, 213]. Interaction metaphors that involve virtual hands are improved by increasing users' control over the hands by affording finger motions, and by providing visual feedback [109, 202]. Furthermore, it bodes well to provide additional feedback through the interacting entities (virtual hands and objects) by using illumination effects [202] and indicating the grasping status [147]. A recent study comparing grasping feedback modalities suggests that users prefer auditory feedback of the grasping status and events over visual feedback [36] for pick and place tasks. Research on grasping and object manipulation in VR suggests that high fidelity grasping using tracked gloves leveraging IMU sensors performs better than a camera vision-based commodity hardware like the

LeapMotion sensor [126]. By grasping objects in VR, it is possible to drag, resize, rotate, and transform objects in space.

Near-field interaction in contemporary augmented reality is achieved primarily through freehand interaction. Contemporary research on interaction techniques in AR seeks to improve and further evolve gesture-based interactions into more physical ones thereby mimicking how interactions take place in the real world [189, 208]. Research has demonstrated that using hands/arms to manipulate objects in AR offers possibilities to engage upper extremities with users who suffer from motor dysfunction [44]. Prior works have sought to understand the intricacies associated with freehand interaction-based grasping in mixed reality and outline difficulties such as object displacements and inaccurate object size estimations as a result [2, 3]. Relatively recent research suggests that users can learn how to interact with and manipulate objects using gestural interaction in a fairly short amount of time in AR [179]. In the aforementioned study, users were able to quickly learn how to directly manipulate virtual objects given how the objects moved with their hands in an intuitive fashion. Just like in VR, the type of means used to support hand interaction can also affect object manipulation performance [49, 100, 171]. The level of naturalness associated with the different gesture-based techniques can also affect how well users can perform simple object manipulation tasks like translation, rotation, and scaling [4]. Results of the aforementioned study suggest that users may choose less natural approaches even though they are less enjoyable because they can be easy to learn. Object manipulation in AR can be facilitated through two natural hand interaction paradigms, namely the metaphoric and isomorphic paradigms each of which is associated with its own pros and cons [196]. In a study comparing the two on a precise fine motor object manipulation task, it was found that users perceive the latter to be more natural, usable, and more accurate for a displacement task [64]. For resizing objects, however, the former outperformed the latter. Users in the study were tasked with moving, rotating, and scaling virtual objects (cubes) using both paradigms. As such near-field object interactions in AR revolve around users' ability to use natural hand gestures and freehand interaction techniques to be able to manipulate objects effectively and efficiently.

Quite a number of studies have looked into comparing selection performance between VR and AR. Along these lines, Pham and Stuerzlinger compared three input devices, namely a mouse, pen, and controller on a pointing task in both the mediums [165]. Their results suggest that a 3D pen was able to outperform VR controllers, and more importantly, that VR was better in terms of pointing performance. They attribute this contrast in performance between the mediums to the difference in display latency characteristics, comfort associated with the HMDs, and the visibility of the real world in AR. Results from another study comparing object selection performance between the two mediums showed that user performance was better in AR than their simulated mirror-based VR condition [188]. It is likely that the difference in a multitude of factors like the visibility of the objects, head-tracking, and occlusion influenced their results. Comparing user performance in an eye-hand coordination training system based Fitts' task in VR and AR, Batmaz et al. show that interactions operationalized through the measures of speed, precision, and accuracy were superior in VR than in AR. Additionally, their results suggest that provisioning haptic feedback does not improve interaction performance in both mediums [14].

There have been a few studies that get into comparing the two mediums, that is VR and AR, on near-field object manipulation tasks in the same setting. In the same vein, work conducted by Botden et al. evaluated two systems (one VR, and the other AR based) in a laparoscopic surgery training scenario [25]. Their results suggest that users performed a suturing task better with the AR system. It was also reported that users found the AR system to be more realistic, useful for training, and have better haptic feedback than the VR system. This being said, their primary goal was to evaluate two competing systems rather than compare the mediums per se. Research investigating the consequences of different visual effects like blurring and fading on a hand-based object manipulation task in both VR and AR suggests that neither the visual effects applied nor the mediums of interaction affect task performance [43]. However, the authors noted what seemed to be an interaction effect between the visual effects applied and the medium of interaction. Another study that compared object manipulation performance in stroke patients using a pick-and-place rehabilitation task showed that users performed better in AR than in VR [99]. It must be noted that the aforesaid study investigated non-immersive monoscopic screen-based VR and projector-based tabletop AR where users interacted with a physical object. The mediums primarily served to display the target object placement areas, implying that participants in the VR condition had the added load of not being able to see their hands when performing the task. Even slight deviations when replicating a real-world environment in the virtual world can affect object manipulation performance in VR [84]. Users in this study performed an object sorting task in either VR or in the real world and it was found that task duration and the duration of eye fixations were more for users in VR than those in the real world. Given a low sample size (4 users per condition), it must be acknowledged that the results may not generalize to other application scenarios. Prior studies measuring speed and accuracy in object placement tasks also demonstrate that when correctly aligned, the virtual hand control input metaphor outperforms a table-bound 3D mouse cursor input device [229]. The authors however note that their task may have been the reasons for such observations. Users in the study could only perform rotation and translation transformations one after the other, thereby not resembling a more natural interaction paradigm. Usually, both those actions can be done simultaneously to home in on the required or rather desired result. Moreover, their hardware can be considered an anachronism with respect to contemporary immersive AR technologies making their results less applicable to current scenarios. From a contemporary standpoint, a recent comparative study investigated the effects of the medium of interaction on 3D object manipulation performance. Users were tasked with performing an object selection and transformation task using both a 3D input device and a mouse for both mediums. It was found that users were able to complete tasks faster in AR than in VR for both the type of input devices used to manipulate the objects [108]. The authors fall back on arguments like increased user engagement in AR and reduced environmental complexity in their VR condition to explain their observed results. It is difficult to say which particular component of the visual stimuli in the AR setting contributed to the outcome.

2.9 Interactions with Distortions, Gains, and Offsets

Similar to the distortions that can come from stereoscopic voxels in 3D displays, space can be distorted in other ways as well, including a discrepancy between the user's physical end effectors and the proxy that represents them. For instance, Batmaz and Stuerzlinger conducted a study comparing movements in the presence or absence of rotational jitter in a motion tracking setting, finding adverse effects of jitter in error rate, movement time, and overall performance at different thresholds of jitter [15]. In a similar manner, Brickler et al. found that there may be a threshold of adding mismatch between movement of a user's physical end effectors and the corresponding virtual selector [28]. Similar to this mismatch, Kohli et al. indicate that adding a spatial discrepancy to the direction of the user's end effectors significantly increases movement time for a given selection, but not in overall performance [103]. In addition to these spatial discrepancies, translational offsets and gains between the physical and virtual end-effectors have been studied as well. Fu et al. conducted a study on the effects of visual and haptic co-location through the lens of Fitts' law, finding that co-location improved accuracy [65]. Similarly, Li et al. found that task performance was improved without offsets and with added kinesthetic feedback [118]].

There are various approaches to implementing hand offsets for redirection purposes in immersive environments, allowing for both uniform and non-uniform application of these offsets. Benda et al. conducted research on fixed hand offsets in six directions and identified significant variations in detection thresholds based on the offset direction [16]. Notably, the negative Z ("close") offset direction exhibited the smallest threshold range of 7.83 cm, while the negative Y ("down") offset direction had the largest threshold range of 13.37 cm. This suggests that both the magnitude and direction of the offset can influence the threshold for offset detection. In a two-dimensional (2D) context, Kohm et al. employed a two-alternative forced choice (2-AFC) method to measure human perception of offset hand placement by applying fixed positional offsets to a virtual hand [104]. On the other hand, gain-based interactions, which involve non-uniform offsets, are also utilized in certain studies as an alternative means of seamlessly redirecting a user's hand in virtual reality [57, 235]. In these cases, the offset tends to increase as the user's hand moves further from their body. Ogawa et al. discovered that factors beyond the offset amounts, such as the realism of virtual avatars, can significantly impact the perception of proprioceptive offsets [156]. It is intriguing how a mere change in self-representation within a virtual environment can influence a user's affordances. Azmandian et al. introduced haptic retargeting as a technique that maps an object in the non-virtual world to multiple objects in the virtual world, presenting various methods for aligning physical and virtual objects, including world manipulation, body manipulation, and a hybrid technique [10]. Their study introduced the body warping technique, which applies translation offsets to achieve the desired hand movements and reach the intended destinations. Hand offsets are often combined with other techniques to accomplish hand redirection effectively, as demonstrated in previous studies [69, 140, 207]. These studies illustrate how a user's action capabilities for reaching can be modified in virtual reality, whether through fixed or non-uniform offsets. Despite the differences in these methods, participants have shown changes in sensitivity and behavior when exposed to these offset presentations.

Chapter 3

The Effect of Virtual End-effector Representations on Near-field Interactions Associated With Dynamic Affordances in Immersive Virtual Reality

The work presented in this chapter was supported in part by the US National Science Foundation (CISE IIS HCC) under Grant No. 2007435. A substantial portion of the contents of this chapter was published in the IEEE Transactions on Visualization and Computer Graphics journal 2023 [220].

3.1 Motivation and Overview

The real world features interactions with different dynamics. In some scenarios, humans interact with components that are relatively static or stationery. Picking up a book, reaching to grab your coffee mug, typing on a keyboard, connecting your charger to the outlet on the wall., are examples of such situations where the objects that the human interacts with remain stationery. In many other scenarios, humans interact with objects that are moving. Catching a moving ball, crossing a gap in traffic, working on objects in an assembly line in a factory, etc., are such examples of where the human is required to time their movements to the movements of the objects in order to be able interact with objects and manipulate them successfully. In fact, in most sports involving a ball, humans need to be able to synchronize their motions to be able to effectively catch, pass, intercept, block, smash, and defend the ball. In the past, virtual reality researchers have investigated the effects of different end-effector representations on the perceptions of aspects like size [155], passability, and depth [55], and the performance associated with simple object manipulation interactions like selections and placements wherein the tasks used involved static components [130]. There is however a gap in our knowledge about how such virtual representations affect our perceptions and interaction capabilities in contexts involving dynamic elements. Such dynamic environments and stimuli exhibit spatio-temporal characteristics that are not constant or rather static. Given the importance of interactions in such contexts, many that are likely to feature in a multitude of virtual reality applications once the technology becomes a more pervasive computing platform, we attempt to investigate how end-effector representations affect the perceptions of affordances that dynamically change over time.

In this contribution, we empirically evaluated how different virtual end-effector representations affect users' performance in near-field object retrieval interactions that involve dynamic affordances. Users were tasked with retrieving a target object from a box for a number of trials while avoiding collisions with its moving doors. We employed a 3 (virtual end-effector representation) X 13 (frequency of moving doors) X 2 (target object size) multi-factorial design, manipulating the input modality and its concomitant virtual end-effector representation as a between-subjects factor across three experimental conditions: (1) Controller (using a controller represented as a virtual controller); (2) Controller-hand (using a controller represented as a virtual hand); (3) Glove (using a hand tracked hi-fidelity glove represented as a virtual hand). The frequency of the moving doors and the size of the targets to be retrieved were manipulated within participants. We go on to discuss the implications of these representations on users' sense of embodiment, detailing scenarios that merit using one representation over the other.

3.2 System Description

3.2.1 Apparatus

The IVE used for this study was built using the Unity 2020.2.2f1 game engine software and was rendered on an HTC Vive Pro HMD using a computer equipped with an Intel i7-8700 processor, 32 GB of RAM, and an NVIDIA RTX 2080 graphics card. The HMD has an FOV of 110°with a frame refresh rate of

Figure 3.1: Schematic representation of the joints and phalanges tracked by the MRTK framework on the Microsoft Hololens 2.

90 Hz. The HTC Vive Pro controller and the Noitom Hi5 VR glove (see figure 3.1) were used to provide the end-effector representations investigated in this study. During pilot testing, the simulation's frame-rate was measured, ensuring that it was stable and approximately equal to the device's maximum refresh rate (90Hz).

3.2.2 Virtual Environment

A spacious, rectangular virtual room was designed for this experiment. The room contained a couch, a rug, potted plants, coffee tables, wall paintings and a desktop workstation in the corner. The room was designed to be commodious in an effort to avoid inducing claustrophobic effects. A virtual wooden box was placed on a virtual table in the center of the room and this box was used to host the object retrieval task described in section 3.3.1. A stationary virtual ball treated as the target object to be retrieved, sat in the center of the floor of the box. A uniformly patterned and non-solid texture was applied on the ball to make its contour salient. Users sat on a wooden chair that was physically co-located with a virtual world replica, facing the front of the box (figure 3.2). The virtual wooden chair was modeled and textured to match the dimensions and look of its real world counterpart.

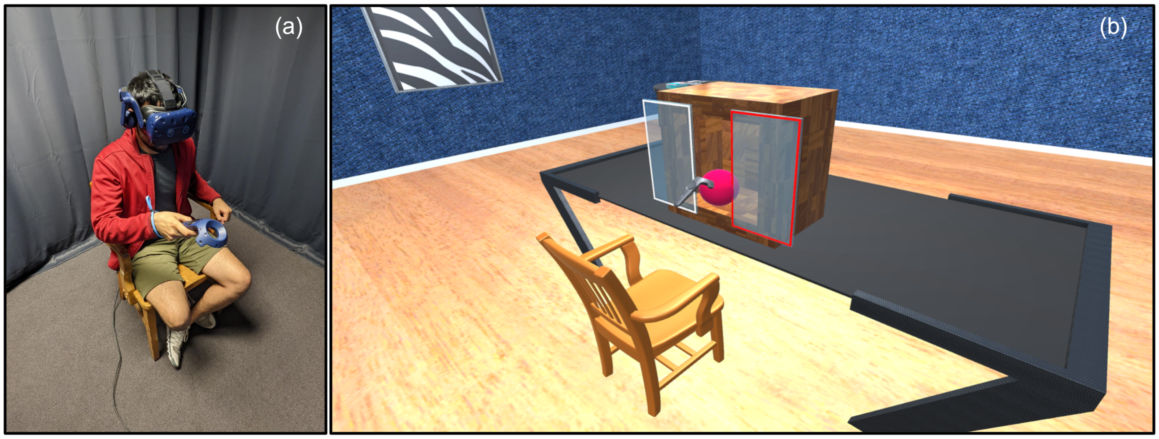


Figure 3.2: Images showing participant performing the task from a third person perspective in (a) the real world, and (b) the virtual world. The highlighted door and ball in (b) denotes a failed trial wherein the participant failed to successfully retrieve the object without collisions.

The box was modeled to be 50cm wide, 35cm deep, and 50cm high, allowing for ample room to retrieve a target from inside. The front of the virtual box featured two identical rectangular, transparent sliding doors made out of glass. The bezels of the doors were designed to be metallic to make the edges of the door appear salient and distinct. The two doors were programmed to simultaneously and symmetrically oscillate along a horizontal axis in a fashion that opened and closed the box periodically. A custom script was programmed to control the periodicity of the doors' oscillations, thus allowing to control the number of times that both the doors slid in and out per minute (referred to as door frequency). A higher frequency implies a larger number of oscillations per minute, which corresponds to a higher speed at which the doors oscillate. When the box is closed, the inner edges of the doors touch, thus closing the aperture from which the target can be retrieved. The box is fully open when the distance between the inner edges of the doors become equal to the width of the box, representing the largest aperture width from which the target can be retrieved. The sliding doors correspond to a dynamic affordance wherein the width of the aperture from which a target can be retrieved, continuously changes at a constant rate. This aperture width increases when the doors slide out and decreases when the doors slide back in. Sounds were not added to the moving doors to ensure that participants performed the task strictly based on visual information of the aperture. Furthermore, pilots revealed that users felt annoyed when a collision sound was periodically produced on the doors coming to the close position when their inner edges touched.

3.2.3 End-Effector Representations

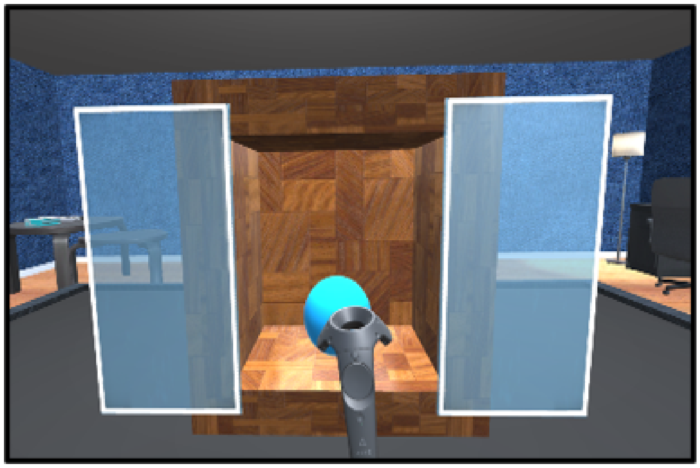
This study investigated three different virtual end-effector representations the specifics of which are described in this section. We use the term end-effector as it encompasses terms that refer to either the controller, the virtual hand or more broadly speaking, the tool with which users interact in the virtual world. The following three representations were investigated in this study.

Controller: This end-effector representation provides users with an identical virtual representation of what they hold in their hands in the real world. The HTC Vive Pro controller is rendered in the virtual world based on the 3D models provided by the Steam VR plugin. This 6 DoF physical handheld controller is collocated with its virtual replica.

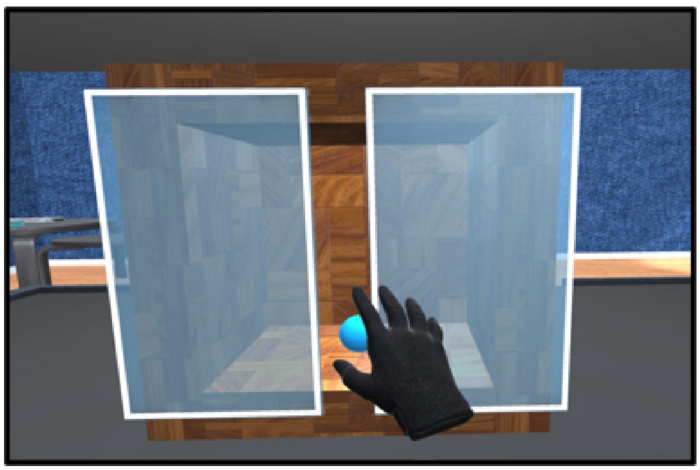
Controller-hand: This end-effector representation provides users with a realistic looking virtual hand when holding an HTC Vive Pro controller in their hands in the real world. The virtual hand models are provided by the Steam VR plugin. The position of the user's thumb on the controller's touchpad is mapped to the position of the virtual hand's thumb. Animations are applied on the virtual hand model depending on the button being pressed, taking into account the position of the thumb. The animation involves a fist clenching grasping gesture.

Glove: This representation provides users with a 6 DoF hand-tracked high fidelity realistic looking virtual hand when wearing the Noitom Hi5 VR glove [153]. The gloves are compatible with the HTC Vive Pro tracking system such that they are fitted with an HTC Vive tracker to track the position and orientation of the glove in the virtual space. The Noitom Hi5 VR Glove delivers wireless, full-finger tracking with a series of IMU sensors that can accurately relay un-occluded motion data in real-time. With this representation, the full-finger tracking allows for poses and gestures to be sensed.

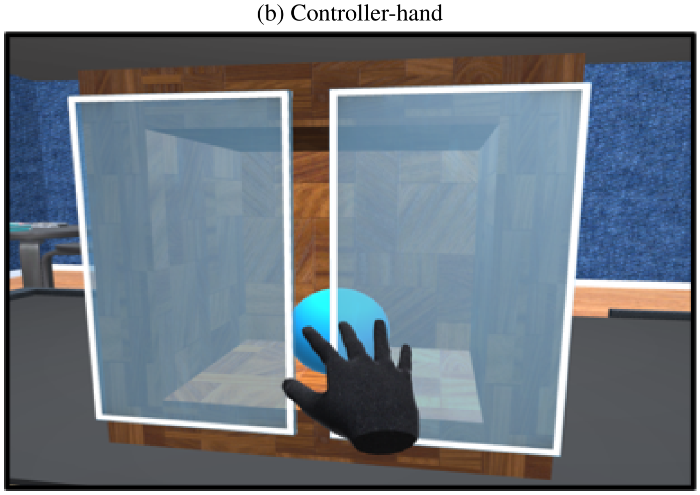
In all three conditions, the end-effectors were tracked using the HTC Vive Pro tracking system. The controller and controller-hand representations were identical to those employed in [130]. In both these representations, the control mechanism was exactly the same. To perform the actions necessary for the task, only the trigger button on the back of the controller, accessible with the index finger, was required. In order to grab the target ball, users would approach it and press the trigger button. With the hi-fidelity glove, the provision of six 9-axis IMU sensors on each finger allowed for high performance positional and rotational tracking of the hand. Users wearing these gloves could grab the target ball by simply clenching their fists. The distance that a representation had to have to successfully grab the ball was the same across all three conditions. In both the controller-hand and glove representations, the material and texture applied on the hand meshes



(a) Controller



(b) Controller-hand



(c) Glove

Figure 3.3: First per perspective of the virtual experience in the different end-effector conditions of the study. The images show the two ball sizes tested, and the doors at different points during their oscillations. sub figures (b) and (c) show the doors sliding in towards the close position while (a) shows the doors sliding out towards the widest open aperture width.

were made identical to ensure consistency between conditions. Additionally, the texture and shaders applied on the virtual hands made the representation appear neutral as though the user was wearing a black glove. This was done in order to prevent any possible consequences on the levels of presence or feelings of eeriness that could arise from choosing realistic human textures that are not gender matched, an effect that prior research has demonstrated [192].

A system evaluation of latency and frame rate in all three conditions was conducted using Niehorster et al.'s method [151]. Ten samples of latency and frame rate for simple translational and rotational movements were measured in all conditions. A high frame rate camera, namely an iPhone 11 Pro camera with a frame refresh rate of 240 fps was mounted on a stand to capture the physical devices (either the controller or a hand when wearing the Noitom Hi5 VR glove) as well as their respective end-effector representations through a monocular view port of the head mounted display (with the lens removed). The device was moved in a straight line (translational), and rotated about the vertical axis (rotational) multiple times, thus capturing several respective latency samples. Using video editing software to analyze the footage, latency was computed based on the difference between the number of frames it took for the time of either movement or rotation of the physical device and the corresponding movement or rotation of the virtual counterpart (end-effector representation) over multiple trials. The analysis revealed that the mean simulation frame rate (measured in Unity 2020.2.2f1 using the Stats gizmo) for the different end-effector representation conditions were as follows: Controller (139Hz), Controller-hand (141Hz), Glove (152Hz). The average (mean) end-to-end latency of the different conditions were as follows: Controller (Positional lag =14.16ms and Orientational lag = 14.58ms), Controller-hand (Positional lag = 14.58ms, Orientational lag=15ms), Glove (Positional lag = 11.25ms, Orientational lag = 10.41ms).

3.3 Experiment

3.3.1 Task

For this experiment, a simple object-retrieval task with a dynamic affordance was conceptualized wherein users had to retrieve a virtual ball from a virtual box for a number of trials. The front of the box featured two identical, transparent sliding doors made out of glass with metallic bezels, and the box was placed on a virtual table in front of the user (figures 3.3 and 3.2). We chose to use glass doors to allow users see the target during the trials. Users were seated on a wooden chair that was physically co-located with a virtual

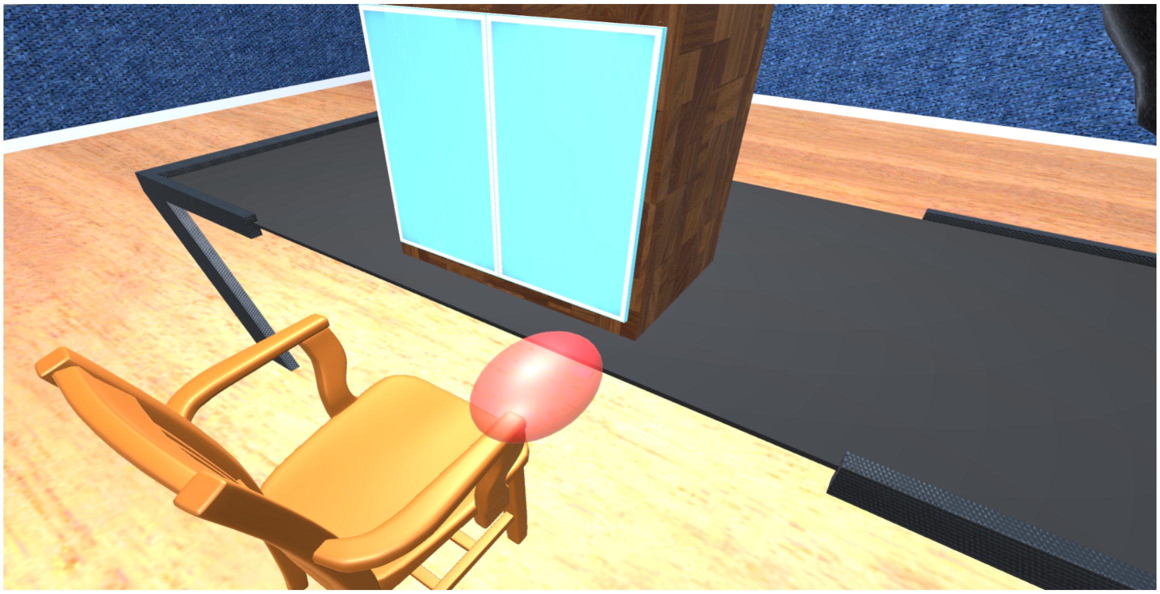


Figure 3.4: Virtual bubble on arm rest of the chair. This red bubble appeared after the commencement of every trial, regardless of the outcome. The red bubble would turn green when users placed their hands on the bubble denoting the start of the next trial. The glass doors were rendered opaque before the start of every trial.

world replica and were tasked with retrieving the ball from inside the box without touching or colliding with either of the doors.

To initiate each trial, users had to bring their arms to rest on the arm rest of the chair (figure 3.4), placing their end-effectors inside a red virtual bubble that sat atop the edge of the arm-rest. Once the end-effector was placed inside the virtual bubble for 2 seconds, the bubble turned green and then disappeared following which the doors turned transparent, marking the start of that trial. This ensured that all participants began every trial with their end-effectors in the same position. In each trial, the doors of the box oscillated symmetrically at a constant speed, thus moving in a way that continuously changed the width of the opening from which the target ball could be retrieved (dynamic affordance). Users were free to take as much time as they required for each trial, and were strongly encouraged to try to be as successful as possible in performing the task.

Each trial resulted in an outcome that was marked with either a success or failure. If the user's virtual end-effector or the target ball collided with either of the doors during a trial, they were provided with both auditory and visual feedback indicating that a collision took place and that their attempt for that trial resulted in a failure. The visual feedback provided during a collision involved the bezels of that door being highlighted in red, the ball being highlighted in a red (figure 3.2), and the doors being frozen in position for 2 seconds.

This was followed by resetting both doors to the close position, re-spawning the target ball at the same target start position, and the glass being rendered opaque, thus preventing the user from viewing the target ball to be retrieved in the next trial. Users would also fail a trial if their end-effector remained completely inside the box while the doors closed, simulating a collision with their forearm despite not rendering it. If the user managed to successfully retrieve the ball, bringing it completely outside the box avoiding any collisions, the ball would immediately disappear and they were provided with a success sound, giving them both visual and auditory feedback that was indicative of a successful trial. After 2 seconds, both doors were reset to the close position and the glass once again turned opaque. On the start of every trial, the doors were rendered transparent and continuously slid in and out. Thus, every trial ended when the user either successfully retrieved the ball or collided with either of the moving doors following which the doors were rendered opaque (figure 3.4) and reset to the close position. It was ensured that the success sound and the sound of a collision were distinctly 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 outcome of that trial.

3.3.2 Study Design

To empirically evaluate how different virtual hand representations affects users' perceptions of dynamic affordances in the near field, we employed a 3 (virtual end-effector representation) X 13 (frequency of moving doors) X 2 (target object size) multi-factorial design, manipulating the input modality and its concomitant virtual end-effector representation as a between-subjects factor across three experimental conditions: (1) Controller (using a controller represented as a virtual controller); (2) Controller-hand (using a controller represented as a virtual hand); (3) Glove (using a hand tracked hi-fidelity glove represented as a virtual hand). Users in each condition performed an object retrieval task for a number of trials during which the size of the target object and the nature or rather rapidity of the dynamic affordance (frequency of moving doors) were manipulated as within-subjects factors. For each experimental condition, participants performed the object retrieval task described in section 3.3.1 for a total of 130 trials. Thirteen different sliding door frequencies were tested in this study ranging from 35 oscillations per minute up to 155 oscillations per minute with standard increments of 10 oscillations per minute between adjacent levels of the frequencies. For each frequency, two different ball sizes categorized as small (diameter=8.61cm) and large (diameter=19cm) were utilized. Each frequency-ball size configuration repeated five times thus making a total of 130 trials (13 Frequencies X 2 Ball sizes X 5 repeats). The order of the all the 130 trials was randomized for each participant.

3.3.3 Measures

Performance - Each trial resulted in an outcome that was marked with either a success or failure in retrieving the target, creating a dichotomous dependent variable. This allowed for the operationalization of performance as a probablistic estimate of successfully retrieving the target object from within the box without any collisions, allowing for classical psychophysical analysis.

Embodiment - Users' level of embodiment towards the virtual end-effector representation was measured using an avatar embodiment survey proposed by Peck et al. [163]. This questionnaire comprises 16 items that load on to four interrelated sub-dimensions (including perceived levels of ownership and agency), which in turn collectively produce a final embodiment score for each user.

Workload - Users' perceived level of workload as a result of the simulation was measured using the NASA TLX questionnaire [74].

Subjective Perceptions (Interviews) - We conducted semi-structured interviews in an effort to qualitatively understand how users perceived the experience. These interviews allowed us to ascertain what strategies they used in performing the task, aspects they found challenging and noteworthy about their end-effectors.

3.3.4 Research Question and Hypotheses

.The overarching aim of this research was to understand if and how end-effector representations affect the perceptions of dynamic affordances in near field interaction. Specifically we were interested in answering the following research question: **"How do end-effector representations affect perceptions of dynamic affordances in near field virtual reality interactions?"** Downstream of this, we were interested in understanding if and how such representations of the virtual end-effector affect perceived levels of embodiment, workload, and presence. User performance was operationalized based on the measure described in section 3.3.3. We developed the following hypotheses that reflect work discussed in sections 2.1, 2.2, and 2.3:

H1: Participants in the controller condition will exhibit the highest success rates.

H2: As the frequency of the moving doors increases, rates of success will decrease.

H3: The target's size will affect the success rate.

H4: Users' performance of the task will improve over trials.

H5: Users in the controller-hand and glove conditions will exhibit higher levels of embodiment than those in the controller condition.

It is expected that the controller condition will outperform the controller-hand condition because the latter involves a discordance between the input modality and its visual representation. With this rationale, one would expect superior performance in the glove condition but given the limited knowledge of the tracking performance associated with this technology in comparison to the HTC Vive controllers, we do not hypothesize effects related to performance in the glove condition. With respect to target size, two competing expositions can be offered. On the one hand, a larger target can be reached faster by virtue of its size making its retrieval easier. On the other hand, the larger target is more likely to collide with the doors during retrieval than a smaller one. For these reasons, we do not develop directional hypotheses with respect to how the target size affects performance. With regards to frequency, it is expected that a higher frequency will yield inferior performance because the frequency represents the speeds at which the doors oscillate.

3.3.5 Participants

A total of 60 right handed participants were recruited for this Institutional Review Board (IRB) approved study, with 20 allotted per condition. Participant ages ranged from 18 to 47 years old (M=22.48, SD=4.69), 35 of whom identified as female. All participants had normal or corrected-to-normal (20/20) vision. Overall, VR Experience did not significantly differ across conditions.

3.3.6 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 VR and video games, experience playing sports. 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, further demonstrating how to use either the controller or the Hi-fidelity glove required to perform the object retrieval task. The experimenter then went on to explain different possible means of failing a trial and how to successfully retrieve the target ball (see section 3.3.1). Participants then donned the HTC Vive Pro HMD (adjusted for their IPD) and performed 5 practice trials to familiarize themselves with the task and its mechanics. The frequencies in the practice trials were 30, 60, 90, 120, 150 oscillations per minute. This ensured that the frequencies presented in the practice phase were different from those tested in the experiment, thus avoiding any potential learning effects. For participants in the Hi-fidelity glove condition, a hand calibration procedure was performed right before the practice phase to calibrate the hand and finger pose tracking associated with the Hi-fidelity glove. After the practice phase, participants then began the experiment performing the task over the 130 trials. Upon completion, participants removed the HMD and filled out the embodiment questionnaire, the NASA TLX questionnaire, and the IGroup Presence questionnaire [87]. 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.

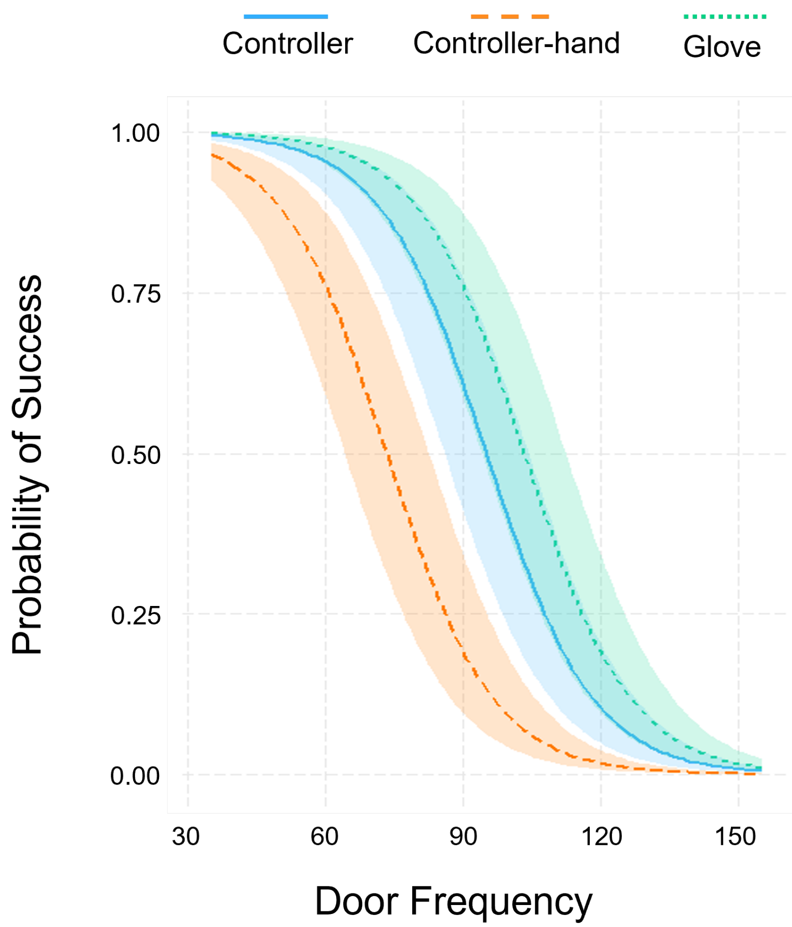


Figure 3.5: Effect of door frequency (oscillations per minute) on success rate, moderated by condition. Shading around each line indicates 95% confidence intervals.

3.4 Results

Since a repeated measures design was used in this experiment, variables had considerable nesting. As each participant completed 130 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 [82]. Since a dichotomous dependent variable, (whether the participant retrieved the ball successfully or not) was used, binary logistic regression was performed. 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 [199].

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 intraclass correlation coefficient was calculated to be 0.155, indicating that approximately 15.5% of the variance in the success rate (whether the participant retrieved the ball successfully or not) 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 binary logistic hierarchical linear model was run to assess the effects of condition, door frequency, ball size and the trial number on participants' success rate. This model with only the main effects (AIC = 4460.8, df = 7) offered a significantly better fit to the data than the null model (AIC = 10013.4, df = 2), χ^2= 5562.5, $p < 0.001$. It explained 82% of the variance in success rate (conditional R^2 = 0.82, marginal R^2 = 0.655).

3.4.1 Performance

3.4.1.1 End-Effector Representation

As expected, there was a significant effect of condition on success rate, χ^2(2, N = 7800) = 23.02, $p < 0.001$, sr^2 = 0.067. Participants were significantly less likely to be successful when in the controller-hand condition (M probability = 0.16, SE = 0.054) as compared to the controller condition (M probability = 0.56, SE = 0.097), z = -3.37, $p = 0.002$, or glove condition (M probability = 0.72, SE = 0.079), z = -4.64, $p < 0.001$. Probability of success was not different when between the controller and glove conditions.

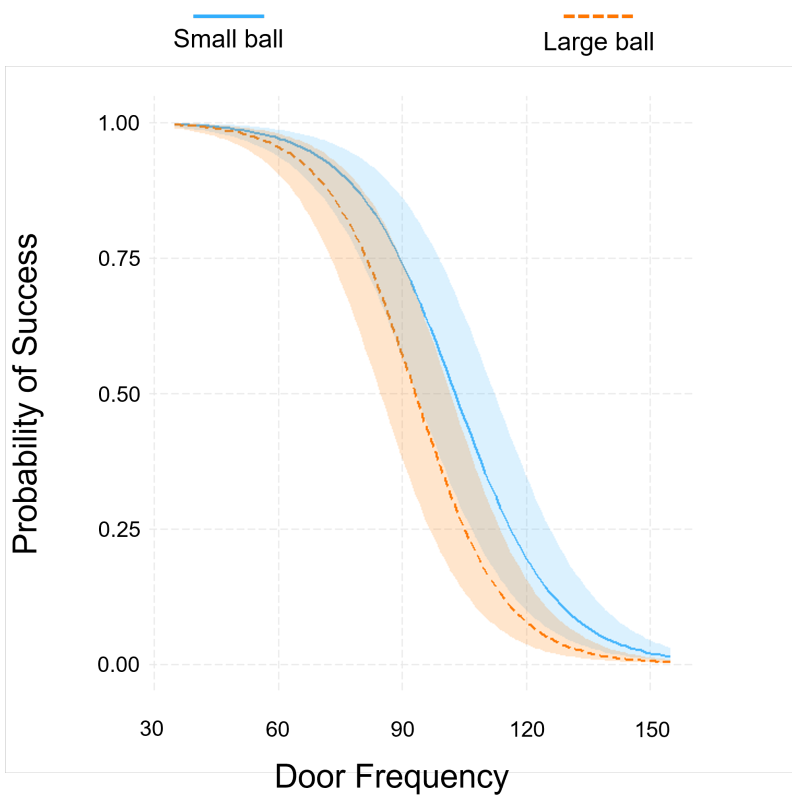


Figure 3.6: Effect of door frequency (oscillations per minute) on success rate, moderated by ball size. Shading around each line indicates 95% confidence intervals

3.4.1.2 Frequency, Target Size and Learning

There was a significant effect of door frequency on success rate, χ^2(1, N = 7800) = 1866.32, $p < 0.001$, sr^2 = 0.60. As the door frequency increased by 1 unit, the odds of successfully retrieving the ball decreased by 0.92. This accounted for 60% of the variance in success rate.

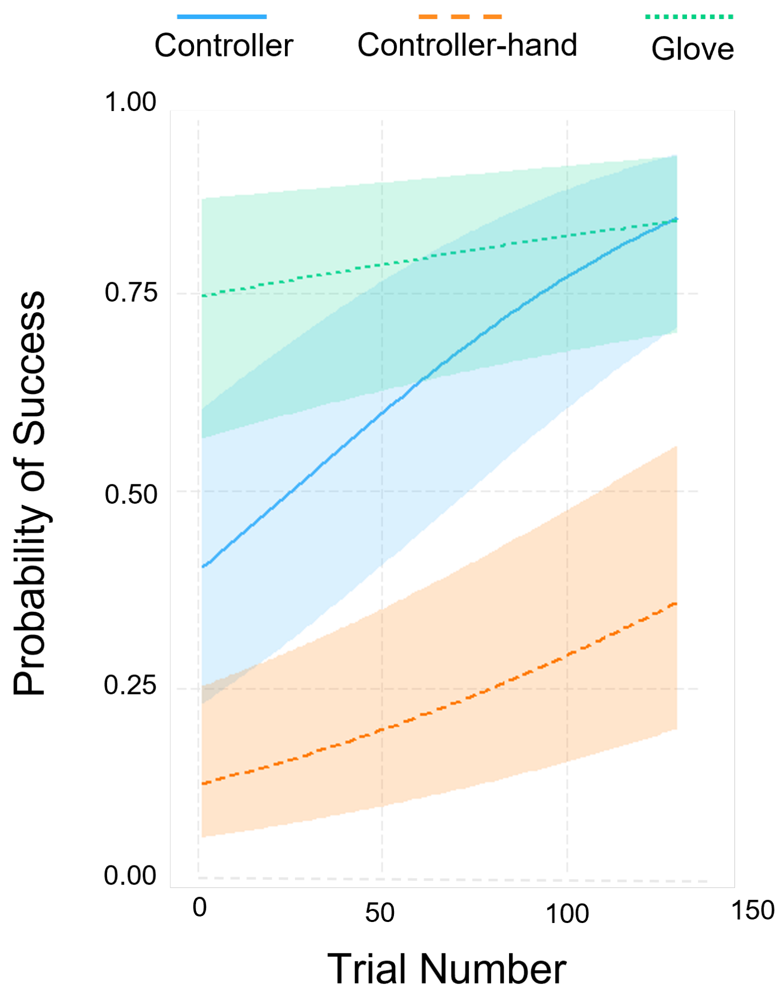


Figure 3.7: Effect of trial number on success rate, moderated by condition. Shading around each line indicates 95% confidence intervals.

Ball size had a significant effect on success rate as expected, $\chi^2(1, N = 7800) = 99.97$, $p < 0.001$, $sr^2 = 0.009$. Participants were significantly more likely to be successful when the ball size was small (M probability = 0.56, SE = 0.06) as compared to when the ball size was lage (M probability = 0.37, SE = 0.05).

The trial number also significantly affected success rate, $\chi^2(1, N = 7800) = 93.04$, $p < 0.001$, $sr^2 = 0.008$. As the trials progressed by 1 unit, the odds of successfully retrieving the ball increased by 1.01. This accounted for 0.8% of the variance.

3.4.1.3 Interaction Effects

Condition was a significant moderator for the effect of door frequency, χ^2(2, N = 7800) = 60.75, $p < 0.001$, sr^2 = 0.02. The levels of condition altered the relationship between door frequency and success rate. As seen in figure 3.5, a test of simple slopes revealed that for each condition, the simple slope for door frequency was negative. The glove condition had a shallower negative slope (B = -0.071, SE = 0.003, odds ratio = 0.93, z = -26.43, $p < 0.001$) as compared to controller-hand condition (B = -0.111, SE = 0.005, odds ratio = 0.90, z = -23.64, $p < 0.001$) and controller condition (B = -0.091, SE = 0.004, odds ratio = 0.91, z = -25.80, $p < 0.001$).

Ball size was also a significant moderator for the effect of door frequency, χ^2(1, N = 7800) = 13.09, $p < 0.001$, sr^2 = 0.002. As seen in figure 3.6, a test of simple slopes revealed that for each ball size, the simple slope for door frequency was negative. The small ball had a shallower negative slope (B = -0.08, SE = 0.002, odds ratio = 0.92, z = -35.40, $p < 0.001$) as compared to the large ball (B = -0.09, SE = 0.003, odds ratio = 0.91, z = -35.60, $p < 0.001$).

Condition was a significant moderator for the effect of trial number, χ^2(2, N = 7800) = 20.89, $p < 0.001$, sr^2 = 0.002. As seen in figure 3.7, a test of simple slopes revealed that for each condition, the simple slope for trial number was positive. The glove condition had a shallower positive slope (B = 0.005, SE = 0.002, odds ratio = 1.01, z = 2.49, $p = 0.01$) as compared to controller-hand condition (B = 0.010, SE = 0.002, odds ratio = 1.01, z = 5.57, $p < 0.001$) and controller condition (B = 0.016, SE = 0.002, odds ratio = 1.02, z = 8.83, $p < 0.001$).

Condition and door frequency were significant moderators for the effect of trial number, χ^2(3, N = 7800) = 12.69, $p = 0.005$, sr^2 = 0.002. As seen in figure 3.8, a test of simple slopes for trial number revealed that, when the door frequency was 57.58 (1 standard deviation below the mean door frequency), the glove condition (B = 0.006, SE = 0.002, odds ratio = 1.01, z = 4.15, $p < 0.001$), the controller-hand condition (B = 0.007, SE = 0.002, odds ratio = 1.01, z = 4.15, $p < 0.001$), and controller condition (B = 0.010, SE = 0.002, odds ratio = 1.01, z = 4.15, $p < 0.001$), had slopes significantly different from zero. When the door frequency was 95 (the mean door frequency), the slope of trial number again significantly differed from zero for the glove condition (B = 0.008, SE = 0.002, odds ratio = 1.01, z = 10.10, $p < 0.001$), the controller-hand condition (B = 0.009, SE = 0.002, odds ratio = 1.01, z = 10.10, $p < 0.001$) and controller condition (B = 0.014, SE = 0.002, odds ratio = 1.01, z = 10.10, $p < 0.001$). Similarly, when the door frequency was 132.42 (1 standard deviation above the mean door frequency), the slope of trial number was again significantly different from zero for the

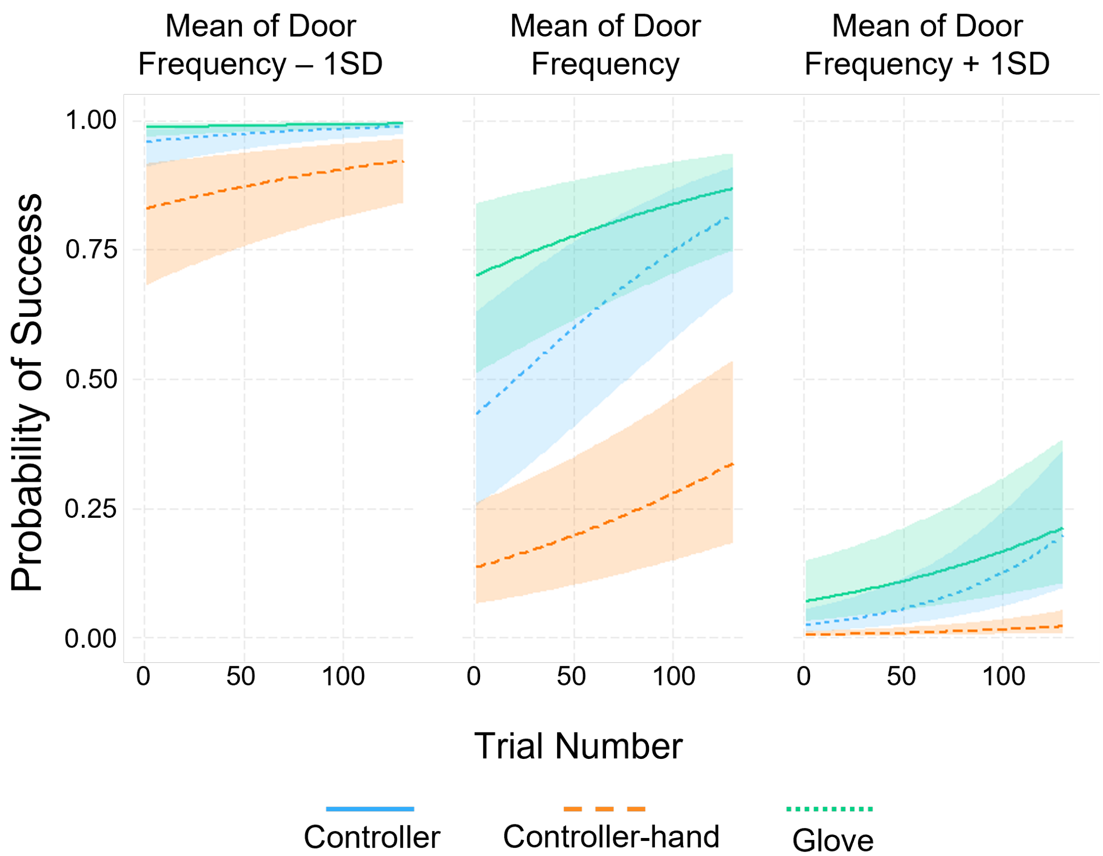


Figure 3.8: Effect of trial number on success rate, moderated by condition at (a) one standard deviation below the mean door frequency, (b) mean door frequency, and (c) one standard deviation above the mean door frequency. Shading around each line indicates 95% confidence intervals.

glove condition (B = 0.010, SE = 0.002, odds ratio = 1.01, z = 7.28, $p < 0.001$), the controller-hand condition (B = 0.011, SE = 0.003, odds ratio = 1.01, z = 7.28, $p < 0.001$) and controller condition (B = 0.018, SE = 0.002, odds ratio = 1.02, z = 7.28, $p < 0.001$).

3.4.2 Overall Embodiment, Ownership, Agency, Workload and Presence

A one-way ANOVA was performed to check the effect of condition on the overall embodiment. The results indicated a significant effect of condition on embodiment, F (2, 57) = 20.04, $p < 0.001$, $\eta_p{}^2$ = 0.41. Post-hoc pairwise comparisons using the Tukey's HSD test showed that embodiment was significantly lower in the controller condition (M = 2.88, SD = 0.92), compared to the controller-hand condition (M = 4.63, SD = 0.64), [t (57) = -5.61, $p < 0.001$], and the glove condition (M = 4.55, SD = 1.30), [t (57) = -5.35, $p < 0.001$]. There was no significant difference in embodiment between the controller-hand condition and the

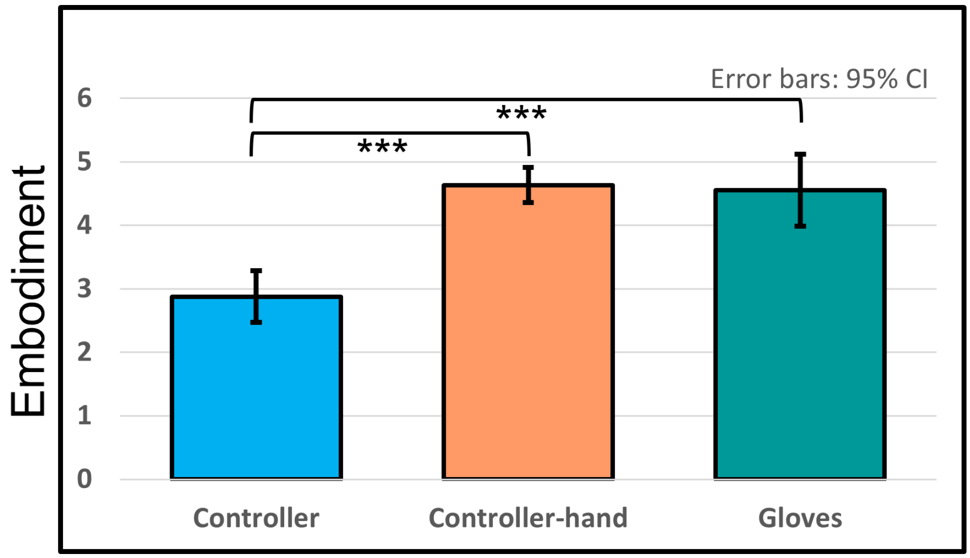


Figure 3.9: Average sense of embodiment scores across the conditions of the experiment. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

glove condition. Figure 3.9 depicts these results.

A one-way ANOVA was performed to check the effect of condition on perceived sense of ownership. The results indicated a significant effect of condition on ownership, F (2, 57) = 23.44, $p < 0.001$, $\eta_p{}^2 = 0.45$. Post-hoc pairwise comparisons using the Tukey's HSD test showed that ownership was significantly lower in the controller condition (M = 3.16, SD = 0.96), compared to the controller-hand condition (M = 5.12, SD = 0.81), [t (57) = -6.25, $p < 0.001$], and the glove condition (M = 4.90, SD = 1.17), [t (57) = -5.55, $p < 0.001$]. There was no significant difference in ownership between the controller-hand condition and the glove condition.

A one-way ANOVA was performed to check the effect of condition on agency. The results indicated a significant effect of condition on agency, F (2, 57) = 20.34, $p < 0.001$, $\eta_p{}^2 = 0.42$. Post-hoc pairwise comparisons using the Tukey's HSD test showed that agency was significantly lower in the controller condition (M = 5.8, SD = 1.79), compared to the controller-hand condition (M = 10.10, SD = 1.71), [t (57) = -6.00, $p < 0.001$], and the glove condition (M = 9.3, SD = 3.05), [t (57) = -4.88, $p < 0.001$]. There was no significant difference in agency between the controller-hand condition and the glove condition.

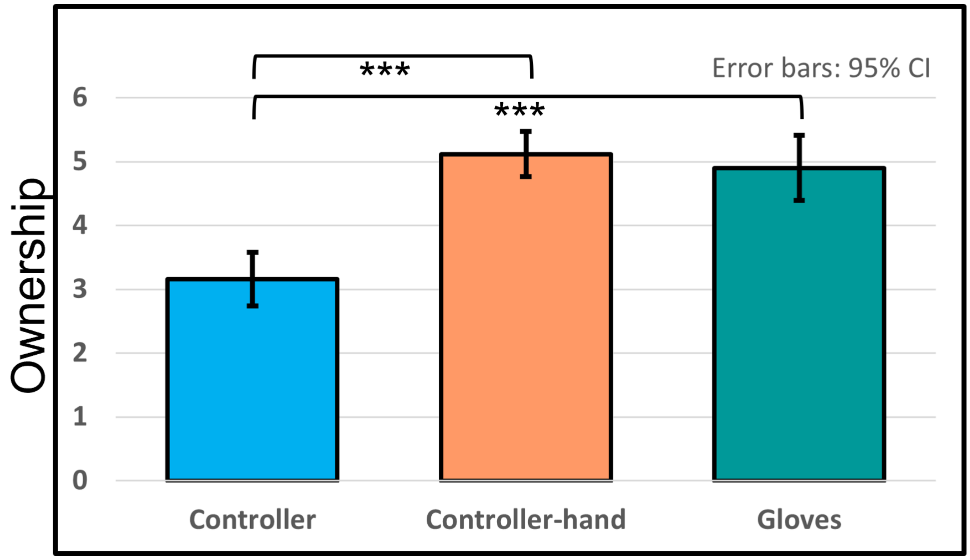


Figure 3.10: Average level of perceived ownership across the conditions. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

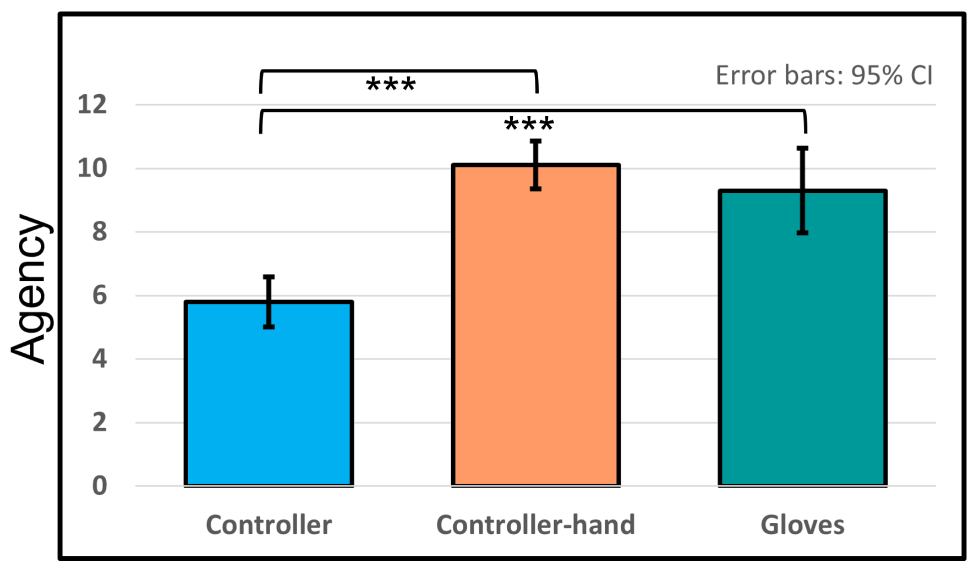


Figure 3.11: Average level of perceived agency across the conditions. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

3.4.3 Workload

A one-way ANOVA was performed to check the effect of condition on the overall perceived workload. The results of the ANOVA indicated a significant effect of condition on workload, F (2, 57) = 4.1401, p =0.02, $\eta_p{}^2 = 0.13$. Post-hoc pairwise comparisons using the Tukey's HSD test showed that workload was significantly higher in the controller-hand condition (M = 7.08, SD = 1.23), compared to the controller condition (M = 5.99, SD = 1.54), [t (57) = -2.56, p = 0.03], and the glove condition (M = 6.05, SD = 1.26), [t (57) = 2.41, p = 0.04]. There was no significant difference in workload between the controller condition and the glove condition. Figure 3.12 depicts these results. There was no significant effect of condition on the measure of presence.

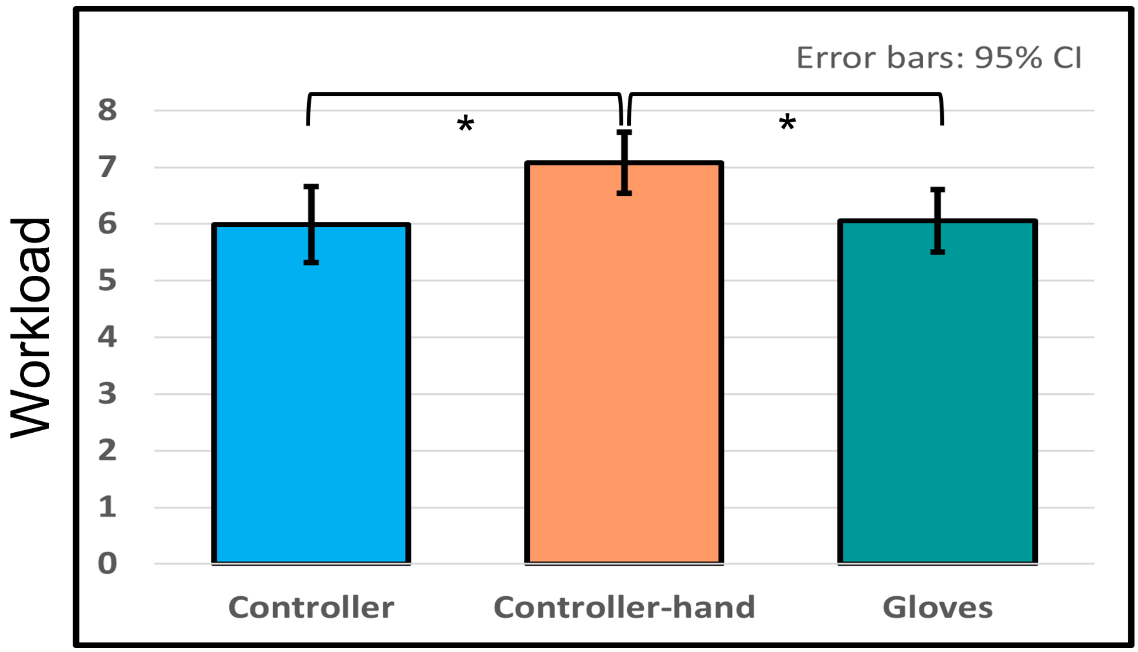


Figure 3.12: Average total workload scores across the different conditions of the experiment. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

3.4.4 Subjective User Experience- Interviews

Data gathered from the semi-structured debriefing interviews shed light on how users perceived the task and experience to be in the virtual world. In relation to the environment, users generally expressed a liking of the virtual room, describing it as a calm, relaxing and visually pleasing room to be seated in. They liked that the virtual chair was co-located with its real world counterpart saying that it reinforced their sense of actually being there. These opinions were expressed by users in all three conditions. With respect to the task, users in general felt that retrieving the target from within the box while avoiding collisions was what made the target retrieval task challenging. The harder ball was deemed easier to grab, but harder to bring out successfully. The smaller ball on the other hand, was described as seemingly easy by many users, given that it was easy to bring it out of the box only when grabbed. Users in the controller-hand condition expressed their liking towards the virtual hand representation. One participant said: *"It's amazing that I can control my thumb using this touchpad, and my fingers using these buttons."*

Another said: *"It feels so real. I like being able to see a hand instead of a joystick."*

Participants in this condition reported that picking up the ball in time was harder especially when the doors were closing or sliding inwards. Many users alluded to the animations of the hand being less natural than how a real hand would grab an object. Users in the glove condition were highly excited about being able to control their fingers accurately. Most of them reported that they had never worn a VR glove before and were happy to get to try it. One of the participants in this condition explicitly said "the tracking with this hand is unbelievable. It feels like it is really my hand in there!" Another user said, "Being able to pick up the ball with my actual hand makes it easier for me". Yet another user said, "I felt like protecting my hand from being hit by the moving doors." Most participants across all three conditions felt like they improved in the task over the trials. However, many participants strongly felt that certain trials, especially the ones with high door "speeds" (the term they used to refer to the frequency), were impossible to be successful in. Another theme that arised out of discussions from the interviews was that some participants, in all the conditions, would time their entry based on a tempo or rather metronome that they calculated. This tempo was based on when the doors would shut on every oscillation. It was interesting to note that the participants that brought this up were all music majors or learned music in the past. There were only 4 participants that reported doing this.

3.5 Discussion

The statistical analyses pertaining to user performance revealed that the end-effector representation had a significant effect on users' ability to successfully retrieve the target from inside the box. The results revealed that there were no differences between the controller and glove conditions, but that the controller-hand condition negatively impacted performance. Figure 3.5 depicts the effects of the door frequency and condition on the probability of successfully retrieving the target without any collisions. The curves resemble psychometric functions wherein differences between curves are typically evident at non extreme stimulus levels (in this case the frequency of oscillating doors). This can be explained in terms of the level of difficulty of the task at the extremities of frequencies tested. At very low frequencies, participants find the task easy to perform and can successfully retrieve the target without any collisions regardless of the end-effector representation. Similarly, at very high frequencies, participants in all conditions fail to avoid collisions. Thus, at these extreme ends of the frequency spectrum, performance does not significantly vary between the conditions. The differences between the conditions on performance manifest in the frequency range that sits in between the extreme frequencies. Along these lines, the curve associated with the controller-hand condition is shifted left in comparison to the other two conditions. Consequently, the Point of Subjective Equality (PSE) in this condition can be seen to be lower than those of the controller and glove conditions. The PSE can be construed as the door frequency that produces a success rate corresponding to equiprobable outcomes (50%) of either success or failure in retrieving the target from within the box without collisions. The lower PSE in the controller-hand condition indicates that the equiproabable likelihood of users failing a trial occurs at a frequency lower than the other two conditions. Taken together, it can be seen that performance was best in the controller and glove conditions and was significantly diminished in the controller-hand condition, supporting hypothesis H1. These findings are in line with prior research that found faster performances in positioning when using a controller represented as itself rather than it being represented as a hand [130]. Furthermore, our findings situate with research showing that the end-effector representation affects the perception of available action possibilities in virtual reality [95], extending the same to scenarios that involve affordances whose spatio-temporal dynamics continuously change over time. Our results hence tend to favor a one-to-one mapping between the input modality used to perform the task and its concomitant end-effector representation, aligning with the idea of a more functionally transparent virtual tool to help improve performance in scenarios involving dynamic affordances.

With respect to the factors affecting the probability of success at a within-participants level, our results showed significant effects of the frequency of the oscillating doors, the size of the target ball to be retrieved and the familiarization of the task over trials. Firstly, as expected, participants were less likely to be successful in retrieving the targets without collisions when the frequency of the doors increased, confirming hypothesis H2. This is understandable given that a higher frequency corresponds to a larger number of oscillations in the same time frame, implying an increased difficulty in performing the task without collisions. This can be inferred from figures 3.5 and 3.6, both of which show declining probabilities of success on increasing door frequencies. Secondly, our results revealed that the smaller ball was associated with a higher probability of successful retrieval without collisions (figure 3.6). This may be because the size of the target to be retrieved dictates the added potential for collisions to occur during object retrieval in such a dynamic affordance task.

Since collisions were based on both the end-effector and the target ball to be retrieved, a smaller ball has lower potential for collision than a larger ball simply by virtue of its size. This was corroborated by users' comments in the debriefing interviews wherein they commented about the larger ball being more difficult to maneuver and bring out through the gap, requiring them to move their hands faster to avoid collisions. Our results hence confirmed hypothesis H3, suggesting an influence of the size of the object to be retrieved in dynamic scenarios involving fine motor control for near field interaction in IVEs. Lastly, we found a learning effect such that participants' ability to successfully retrieve the target for a given frequency of the doors, improved over trials, confirming hypothesis H4. However, this learning effect was more pronounced in the controller and glove conditions than the controller-hand condition as evinced in figure 3.7. This seems to indicate a diminished ability to calibrate to the representation when there is a discordant mapping between the input modality used and its end-effector representation. This counters results obtained in another study showing that adaptation to avatars is fairly fast in VR [95]. In contrast to their work, our study investigated a dynamic context, requiring users to synchronize their movements to an aperture whose width continuously changed over time. This is a noteworthy finding given the growing number of applications that resort to representing users' end-effectors as hands in attempting to increase the perceived enjoyment and realism of the experience. Moreover, with VR training simulations increasingly featuring moving parts like machinery, equipment, etc., it deserves noting that the type of end-effector representation influences users' efficacy in calibrating their performance over time.

Analysis on workloads experienced by users revealed that the controller-hand condition produced significantly higher levels of workload than both the controller and glove conditions. With respect to the perceived levels of embodiment towards the end-effector representation, the results of the analysis revealed that users in the controller-hand and glove conditions reported significantly higher embodiment scores than those in the controller condition, supporting hypothesis H5. Understandably, both conditions that represented the end-effector as a virtual hand scored significantly higher on the construct of embodiment than when the end-effector was represented as the controller itself, aligning with findings obtained from prior research [130]. Given the findings regarding users' performance in the task, these results on workload and embodiment are interesting. Representing the end-effector as a virtual hand when using a controller seems to increase embodiment, but this alteration of the end-effector representation tends to come at the cost of performance in addition to an increased workload in scenarios involving dynamic affordances.

Taken together, it seems appropriate for VR developers to consider the target requirements of an application when deciding on how to represent users' end-effectors in the experience. It also seems important to understand what consequences the representations of these end-effectors have on different aspects of the experience. Along these lines, when higher levels of performance is desired, it bodes well to represent the end-effector as a virtual replica of the input modality used to perform the task, ensuring a concordant mapping between the input modality and the end-effector representation. On the other hand, when the application requires higher levels of embodiment, it seems to be favorable to represent the end-effector as a realistic hand. Choosing to represent the end-effector as something different from the input modality seems to increase the workload of the users, something important for VR system designers to consider, and for researchers to investigate. It hence stands to reason that the type of virtual end-effector representation displayed to the user, and its mapping with the input modality used to provide that representation, togehter strongly influence how well users are able to perform tasks that involve dynamic affordances in the form of threats that are not static. The representations also inherently determine to what extent users embody such virtual representations.

3.6 Limitations

In this study, the dynamic affordance in each trial was created by doors that oscillated periodically at a given frequency. This results in an affordance that is rather predictable in nature because the speed of the doors (frequency of their oscillations) does not change in a trial. For example, when doors open and close in a variable, and thus less predictable, manner, participants' judgments of their ability to locomote through those doors becomes less reliable [133]. Given that dynamic affordances need not necessarily involve periodic, symmetric, and predictable movements, it must be noted that the findings from this study may be limited to such scenarios rather than more complex dynamical systems. The experimental design employed in this work did not include a fourth condition with a data glove represented as a controller and this can be considered as another limitation of this work. While such a condition was considered, the lack of applicability and generalizability of the potential results generated by this condition dissuaded the pursuit of this investigation.

3.7 Conclusion and Future Work

In conducting this study, we empirically evaluated how different virtual end-effector representations affect users' perceptions of affordances with dynamic characteristics using an object retrieval task in the near field. Users were tasked with retrieving a target from a box for a number of trials while avoiding collisions with both of its moving or sliding doors. We employed a between-subjects study design manipulating the end-effector representation across three experimental conditions, namely, Controller (using a physical controller represented as a virtual controller), Controller-hand (using the physical controller represented as a virtual hand), and Glove (using a hand tracked hi-fidelity glove represented as a virtual hand). Results of this study demonstrate that users in the controller-hand condition performed worse than users in the other conditions. Furthermore, users in this condition exhibited a diminished ability to calibrate their performance with respect to retrieving the target while avoiding obstacles, over trials. We find that representing the end-effector as a hand tends to increase users' subjective sense of embodiment, but can also come at the cost of performance, or an increased workload. This leads us to the conclusion that a discordant mapping between the input modality used to support an interaction and its concomitant visual end-effector representation, can cause detriments to performance especially in contexts involving dynamic affordances and fine-motor perception action coordination. It logically follows that virtual reality system developers, application designers, and researchers should carefully consider what the priorities and requirements of a target application are when assigning visual end-effector representations for users to embody. Along these lines, it bodes well for applications requiring high degrees of performance in near field object retrieval interactions with dynamic environmental characteristics to strive for concordance in mapping between the visual representation of the end-effector and the input modality used to support interactions.

The conclusions drawn from conducting this study opens up the floor to many interesting avenues of future investigations. In this line of work, we wish to further investigate how such findings are affected by virtual end-effector representations that fall at different points in the appearance fidelity continuum. This could involve representations that range from abstract representations like capsules, hooks, and graspers to more photo-realistic representations of human-like hands. Further, we wish to explore how manipulations to the nature of the dynamic affordance affects performance of near field interactions. Specifically, dynamic affordances that do not involve periodic oscillations but are rather more unpredictable are of specific interest. We also wish to understand the effects of moving targets in tasks like these wherein the target to be retrieved moves within a volume in the environment over time. Rigorous empirical evaluations conducted in such areas will provide more insights into the effects of virtual end-effector representations on near field interactions in virtual reality experiences. Based on findings obtained in this study, we encourage researchers to further explore what other trade-offs can manifest as a result of discordant mappings between user representations and input modalities used to support interactions in the immersive virtual reality experience.

Chapter 4

The Effects of Avatarizing Users' End Effectors in Near-field Object Retrieval Interactions in Augmented Reality

The work presented in this chapter was supported in part by the US National Science Foundation (CISE IIS HCC) under Grant No. 2007435. A substantial portion of the contents of this chapter was published in the IEEE Transactions on Visualization and Computer Graphics journal 2023 [219].

4.1 Motivation and Overview

Augmented reality technology is steadily growing with scientists and technologists pushing for its widespread adoption. Unlike virtual reality, augmented reality, at least when achieved or rather realized using specialized optical see-through displays, allows for users to continue to be able to see the real world by registering, augmenting, and further displaying virtual components in the real world. Users wearing such head mounted displays like the Google Glass, Magic Leap, and Microsoft Hololens can interact with virtual artifacts (holograms) in the real world, selecting, placing, and manipulating content through space simply by using free hand gestures. Given that the real world surrounding us remains visible when considering augmented reality as a medium for interaction, one may wonder as to whether there is a need for, or whether there are any benefits to avatarization - provisioning users with a virtual representation of themselves or their end-effectors.

The import of this idea can be captured by the following question: "Why give me a self-avatar when I can in fact see my actual body?"! With this presupposition in mind, evidence on the other hand suggests that there are potential benefits in doing this. Along these lines, muscular self avatars in AR have been shown to improve physical performance as a consequence of the Proteus effect [160]. Similarly, augmented virtual arms have been shown to facilitate interactions with objects in the far field that are connected with the AR system, and this too without breaking users' sense of embodiment [62]. Moreover, self-avatarization can be used to augment arms and other extremities for people that may be missing limbs or may have suffered a stroke. While such examples graciously lend support for self-avatarization in AR, it remains to be seen if and how this affects interactions with virtual entities from a performance related standpoint. When the interactions with holographic content depend on the AR system responsible for tracking a user's hand and gestures, and further performing resultant actions to the virtual content based on these tracked gestures, an argument can be made that it may be beneficial to render a visualization of the tracked hand despite one's real hand still remaining visible to them. This becomes a more sound argument when the interactions being performed require fine motor control and precision in perception-action coordination from the user's end. Medical and training applications in AR are two settings that immediately spring up as examples of where the necessity of precision associated with object manipulation aptly complements this idea of visualizing one's end-effector. Furthermore, with the majority of AR applications involving some form of near field interactions with virtual content using tracked hand gestures, it is imperative to understand how self avatarization plays out in such contexts. Given the dearth literature on this topic, a sequential research approach is favorable wherein the potential benefits of avatarization in augmented reality, if any from an interaction standpoint, is investigated first after which, manipulations to the extent of avatarization, anthropometric, and anthropomorphic fidelities of the avatars can be studied. In this work, we focus on the former in an effort to guide our future research directions that delve into the latter.

Specifically, this contribution presents an experimental investigation that empirically evaluates if the presence and further visual appearance, of an augmented virtual hand overlaid over users' tracked hands, affects the performance of a fine motor object retrieval and collision avoidance task in the near field. We utilized a 3 (augmented hand representation) X 2 (density of targets) X 2 (object size) X 2 (virtual light intensity) multi-factorial design, manipulating the augmented 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 end-effector 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

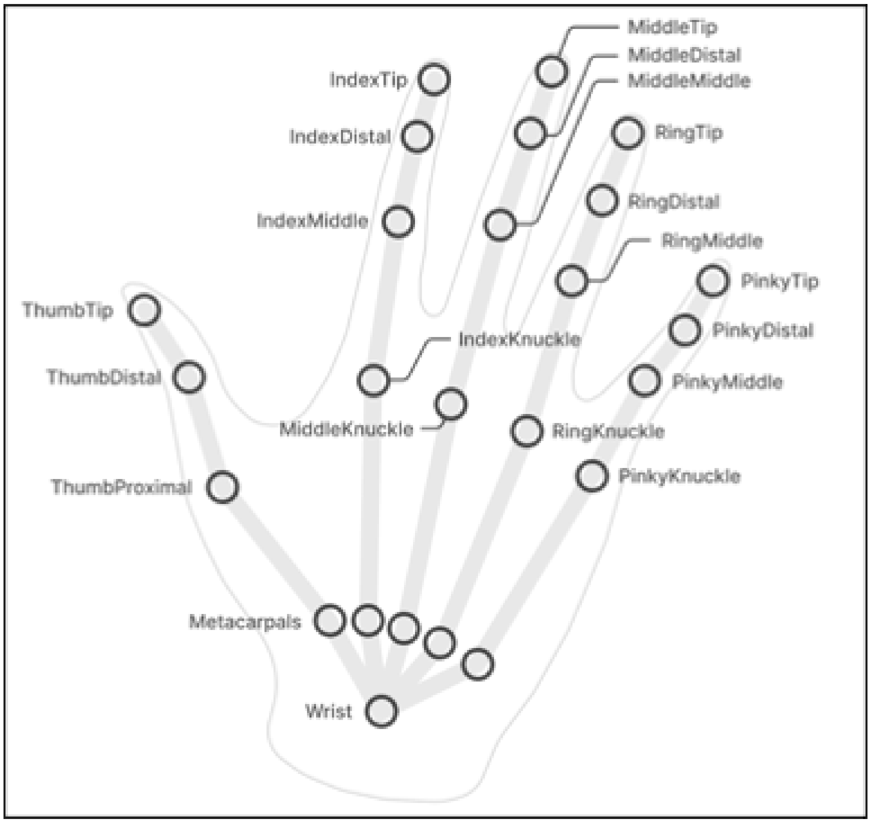


Figure 4.1: Schematic representation of the joints and phalanges tracked by the MRTK framework on the Microsoft Hololens 2.

a collision avoidance based 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. We go on to discuss the results of our investigations and implications of self-avatarization in augmented reality achieved using optical see-through displays.

4.2 System Description

4.2.1 Apparatus

The virtual holograms used for this study was built using the Unity 2020.2.2f1 game engine software and was rendered on an Microsoft Hololens 2 optical-see-through HMD using a computer equipped with an Intel i7-8700 processor, 32 GB of RAM, and an NVIDIA RTX 2080 graphics card. The Hololens 2 HMD 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 which were used in this study. The table top was covered with a black cloth to facilitate better viewing experiences of the augmented holograms. During pilot testing, the simulation's frame-rate was measured, ensuring that it was stable and approximately equal to the device's maximum refresh rate (60Hz).

4.2.2 Virtual Components (Holograms)

A virtual wooden box (hologram) was used to host the object retrieval task described in section 4.3.1. 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. The sky-box associated with the camera was set to a transparent background to make the holograms blend into the real world. 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 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 4.3.1) 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 participants 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.

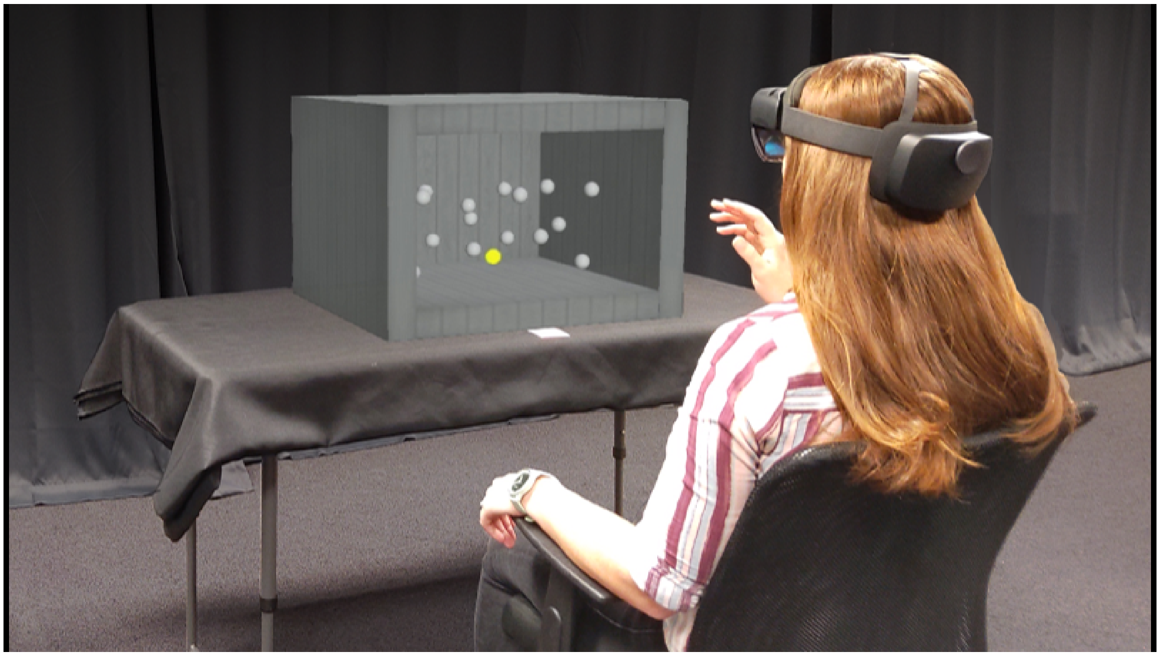


Figure 4.2: Third person perspective of participant performing the object retrieval task. The holographic box is registered and augmented on top of the table.

4.2.3 Hand Representations

This study investigated two different augmented hand avatar representations the specifics of which are described in this section. Note that 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. These representations are depicted in figure 4.4

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 optical see-through display based augmented reality, users can see their own real hands and base their interactions on the same.

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 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 though 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 [192]. 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 hand avatarization conditions (iconic and realistic augmented hand) 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 users actual (physical) hand on top of the table. This average computed positional 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 [151]. Ten samples of latency and frame rate for simple translational and rotational movements were measured in all conditions. A high frame rate iPhone 11 camera (240 fps) was mounted on a stand to capture

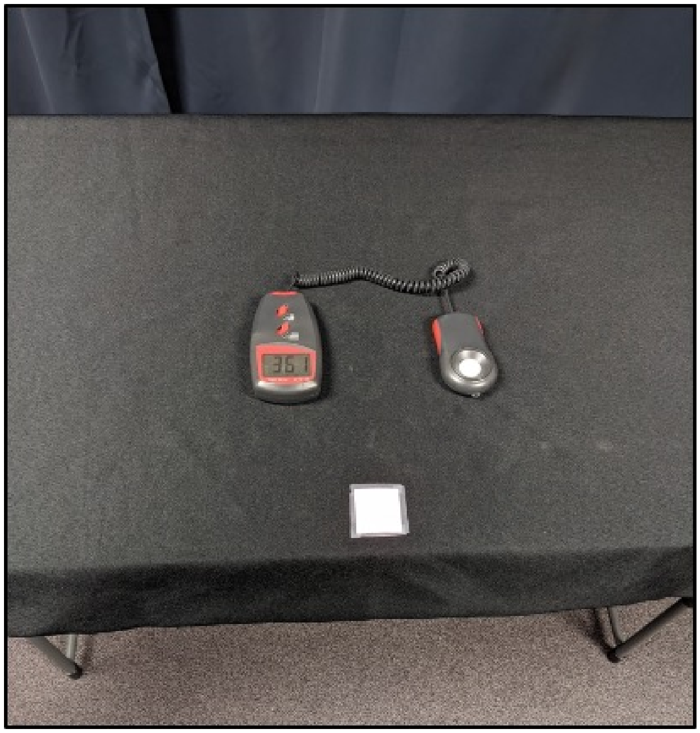


Figure 4.3: Light meter used to measure intensity of light in the real world.

videos of a user's hand as well as the augmented end-effector representations through the view port of the HMD (with the visor removed). The user's hand was moved in a straight line (translational), and rotated about the vertical axis (rotational) multiple times, thus capturing several respective latency samples. Using video editing software to analyze the footage, latency was computed based on the difference between the number of frames it took for the time of movement/rotation of the hand and the corresponding movement/rotation of the virtual counterpart over multiple trials. The analysis revealed that the mean frame rate for the different conditions (sampled in the Hololens 2 using the diagnostics profiler) was measured and found to be 60Hz (refresh rate of the display) 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). These values are observable given that the tracked hand's hologram is behind the real hand when moving with other studies suggesting the same [201].

4.2.4 Virtual Light Intensity

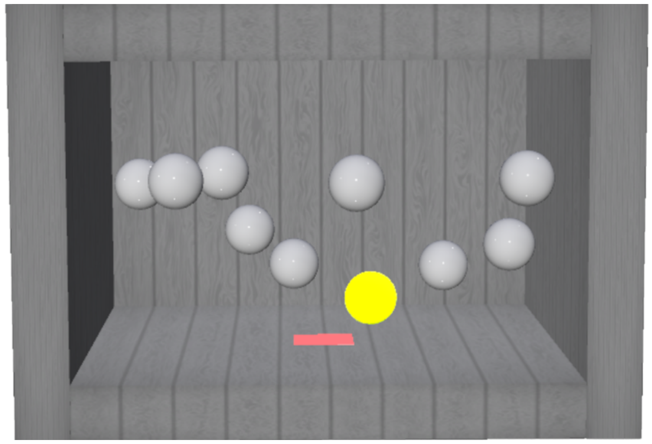
The intensity of virtual light (applied to virtual components) in optical see-through AR can affect the perceived visibility of the real body [224]. With this in mind, we investigated two different levels of virtual light intensity, aiming to study how the resultant visibility of one's real hand impacts interactions in the presence and absence of augmented avatars. The two levels of the light intensity investigated as 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 also measured using a procured light meter [53], and it was ensured that the lighting conditions in the room was kept constant for all participants throughout the course of this study (figure 4.3).

4.3 Experiment

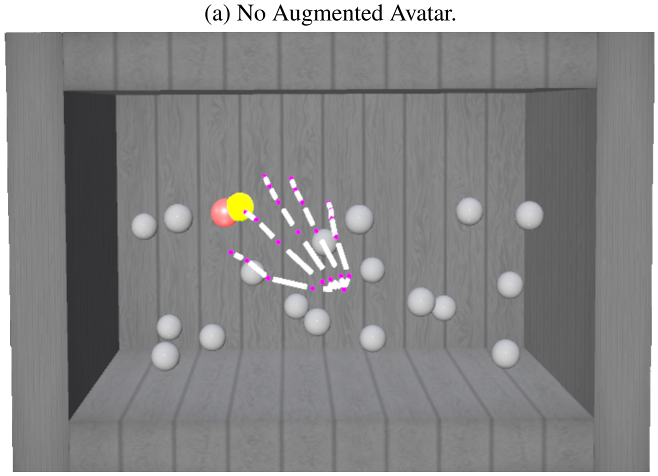
4.3.1 Task

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 virtual box for a number of trials. 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 [209].

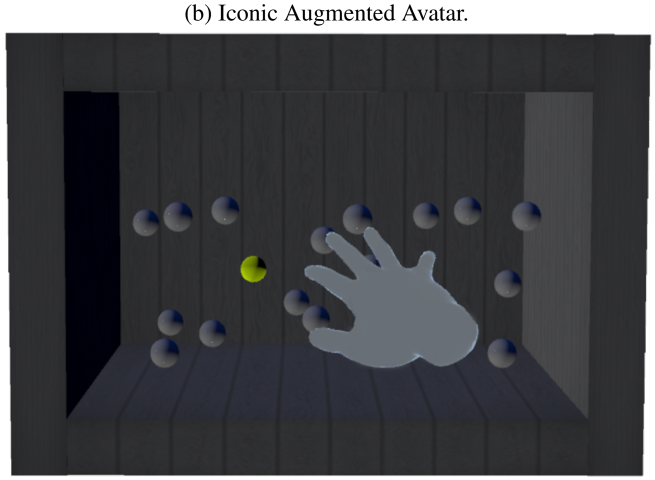
At the start of every trial, the box with its interaction volume was empty. To initiate each trial, users had to push down a virtual button augmented in front of the box. When pressed, the button would disappear and the crowd of balls from which the target had to be retrieved would appear inside the 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,



(a) No Augmented Avatar.



(b) Iconic Augmented Avatar.



(c) Realistic Augmented Avatar.

Figure 4.4: 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.

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 the user managed to successfully retrieve the ball from inside the box, bringing it completely outside the interaction volume, the target and non-target balls would immediately disappear and the user was 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.

4.3.2 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 4.3.1 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.

4.3.3 Measures

Performance - The total number of collision events that occurred in each trial was used an operational measure of performance. This measure 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 4.1. 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. This allowed for the operationalization of performance as the total number of collision events associated with retrieving the target from within the box.

Perceived Real Hand Visibility - At the end of each block, users reported how visible they perceived their actual hands (only the hand; excluding the 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 through the display.

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 [116].

4.3.4 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 optical see-through augmented reality?** 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 4.3.3. Based on this overarching research question, we developed the following hypotheses that reflect work discussed in section 2.2, 2.5, 2.6, and 2.7:

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

H2: A higher obstacle density configuration will be associated with more collision events.

H3: The size of the obstacles will affect performance.

H4: Users' 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.

4.3.5 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 of whom identified as female, 21 of whom identified as male, and the rest as non-conforming. All participants had normal/corrected-to-normal (20/20) vision. Overall, VR and AR experience did not significantly differ across conditions.

4.3.6 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 4.3.1). 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 avoiding any potential learning effects. 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 [116]. 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.

4.4 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 [82]. 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 ϕ is greater than 1 (overdispersion), a negative binomial regression can be used [47]. In this data, since ϕ 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 [199]. 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 intraclass correlation coefficient 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), χ^2 = 1531, $p < 0.001$. It explained 65.6% of the variance in total collisions (conditional R^2 = 0.66, marginal

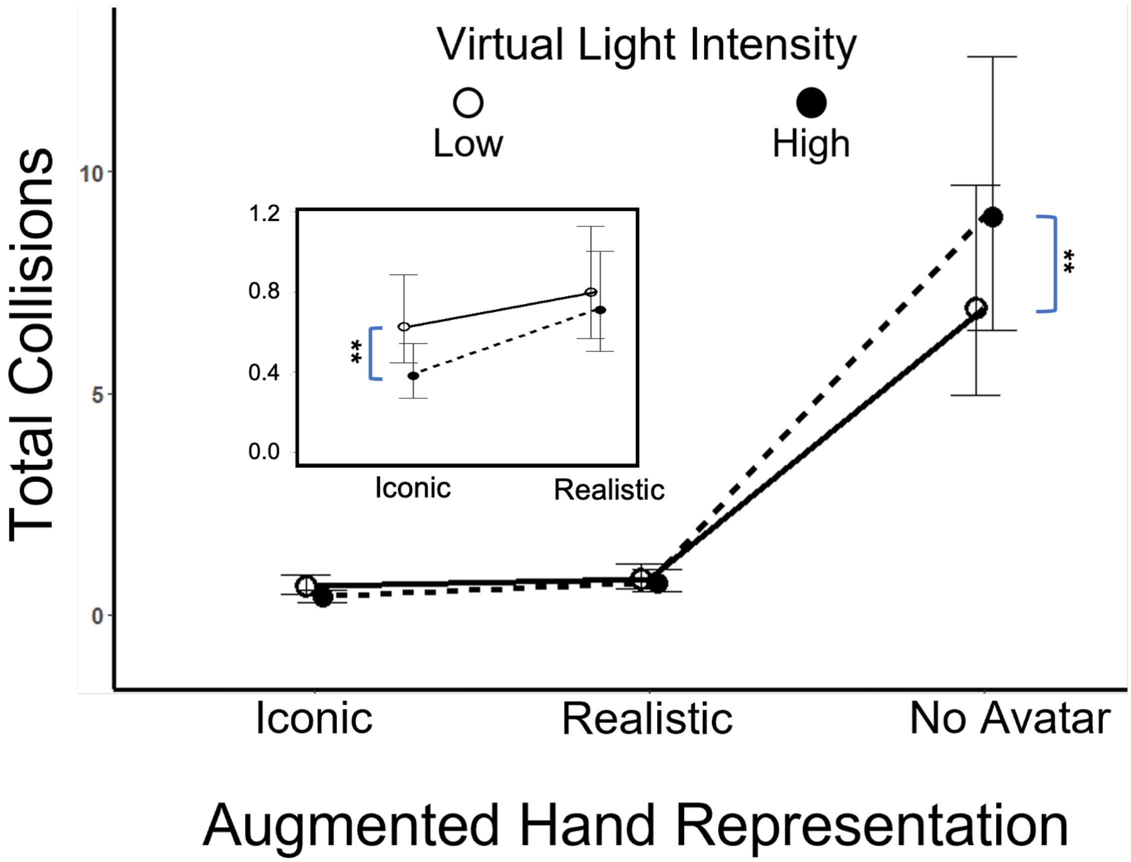


Figure 4.5: Interaction between the avatar condition and real hand visibility. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

R^2 = 0.57).

4.4.1 Performance

4.4.1.1 Augmented Hand Representation

As expected, there was a significant effect of avatar condition on total collisions, χ^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, $p < 0.001$, 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.

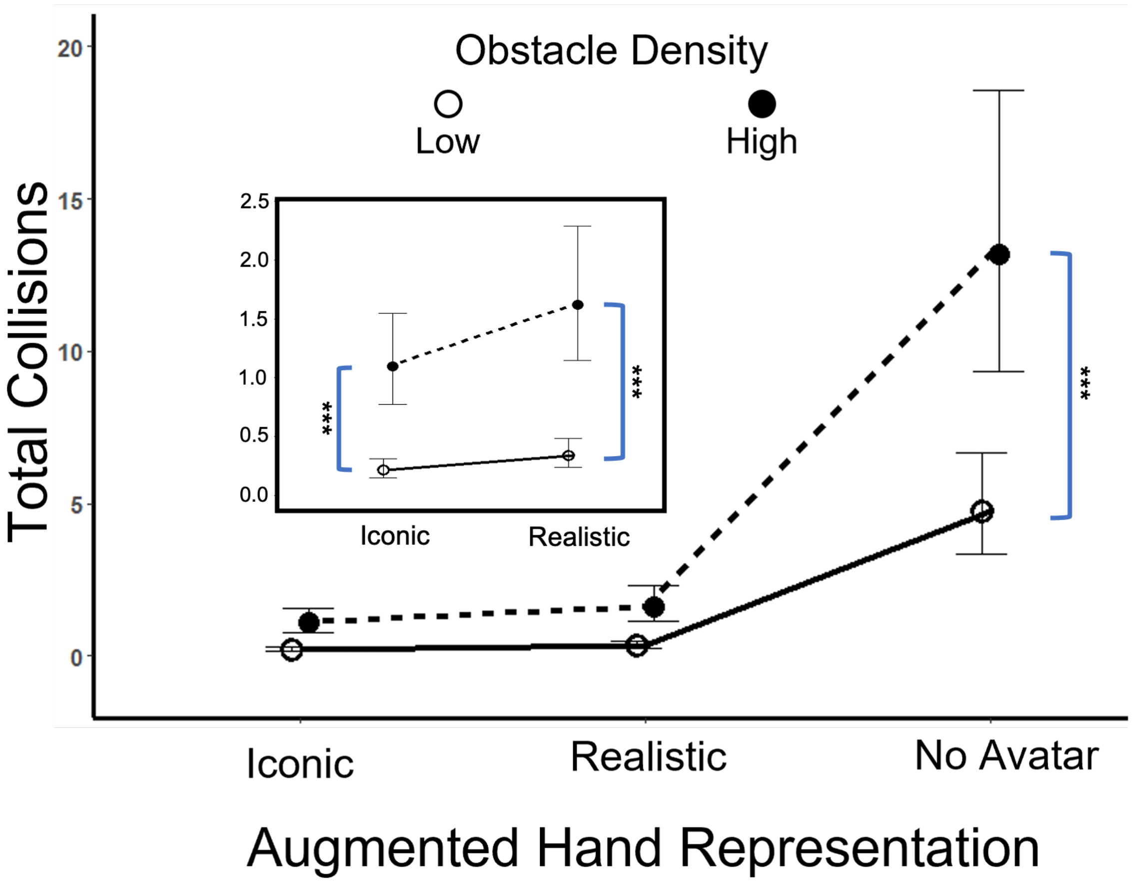


Figure 4.6: . Interaction between the avatar condition and obstacle density. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

4.4.1.2 Obstacle Density, Size, Depth and Virtual Light Intensity

There was a significant effect of obstacle density on total collisions, χ^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 obstacle density (M = -0.31, SE = 0.10).

Obstacle size also had a significant effect on total collisions, χ^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, χ^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.

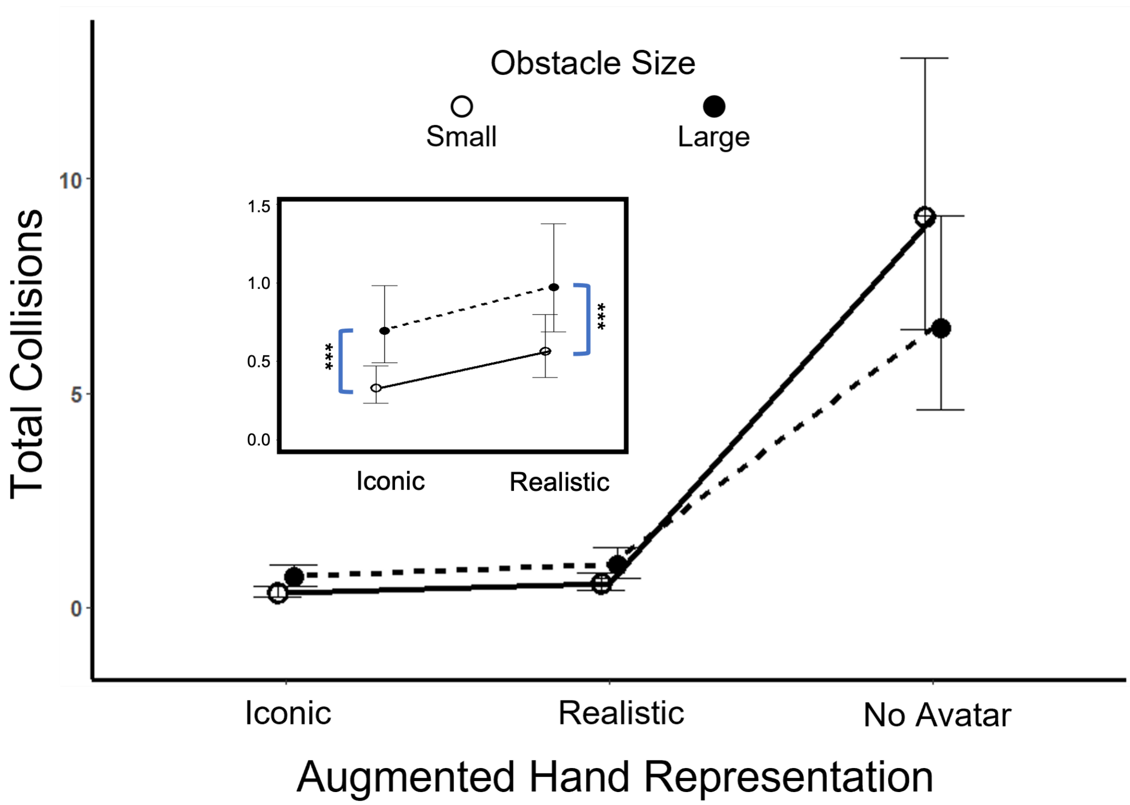


Figure 4.7: Interaction between the avatar condition and obstacle obstacle size. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

4.4.1.3 Interaction Effects

There was a significant interaction between avatar condition and virtual light intensity, χ^2 (2, N = 8160) = 34.14, $p < 0.001$, $sr^2 = 0.005$ (figure 4.5). 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.

There was a significant interaction between avatar condition and obstacle density, χ^2 (2, N = 8160) = 27.34, $p < 0.001$, $sr^2 = 0.005$ (figure 4.6). When testing simple effects, for participants in the Iconic avatar condition, total collisions were significantly more when there was higher density of obstacles (M = 1.09, SE = 0.19) as compared to when there was lower density of obstacles (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, $p < 0.001$.

There was a significant interaction between avatar condition and obstacle size, χ^2 (2, N = 8160) = 84.01, $p < 0.001$, sr^2 = 0.008 (figure 4.7). 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.

4.4.2 Perceived Real Hand Visibility

A Mixed 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 interaction 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 (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 the Realistic (M=4.0, SD=1.45, $p < 0.001$) avatar conditions. Inspecting

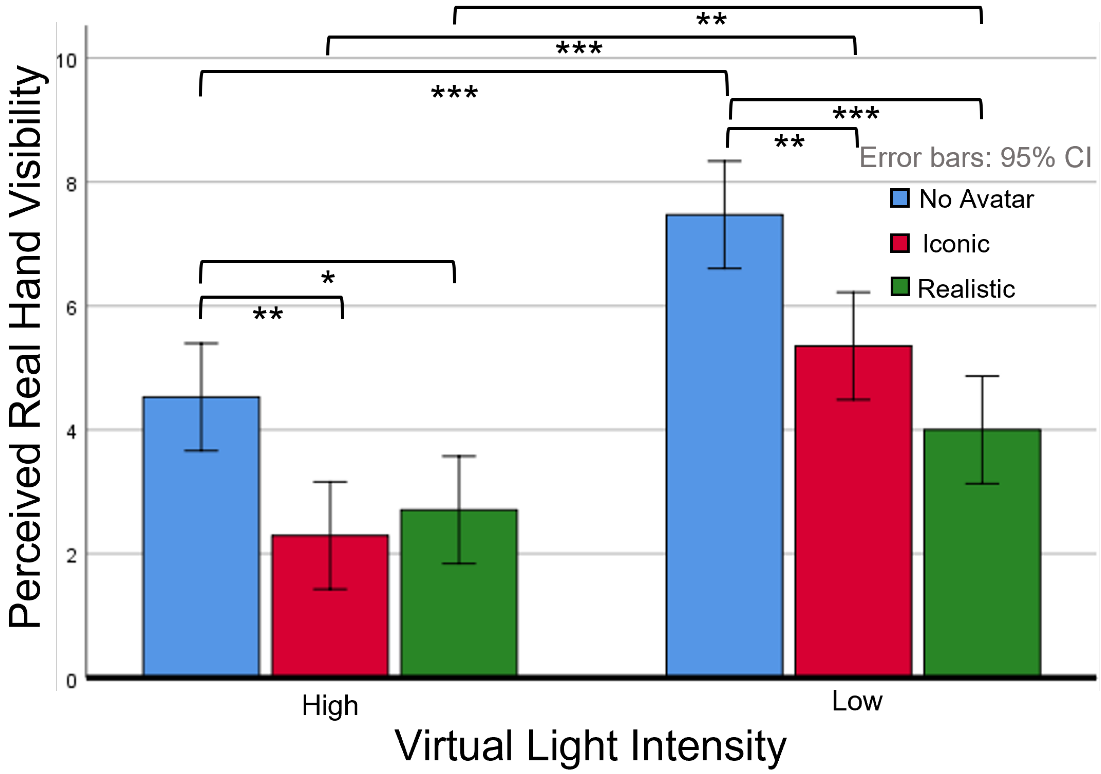


Figure 4.8: Effect of virtual light intensity on perceived real hand visibility moderated by condition. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

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. Figure 4.8 depicts these results.

4.4.3 Perceived Workload

A one-way ANOVA was performed to check the effect of condition on the overall perceived workload. The results indicated a significant effect of condition on workload, F (2, 48) = 3.70, $p = 0.032$, $\eta_p{}^2 = 0.134$. Post-hoc pairwise comparisons using the Tukey's HSD test showed that workload was significantly higher in the No-Avatar condition (M = 5.99, SD = 1.61), compared to the Iconic avatar condition (M = 4.823, SD = 1.09), $p = 0.027$), and not different from Mesh avatar condition (M = 5.59, SD = 1.03, $p = 0.629$). There were no other significant differences in perceived workload between the conditions.

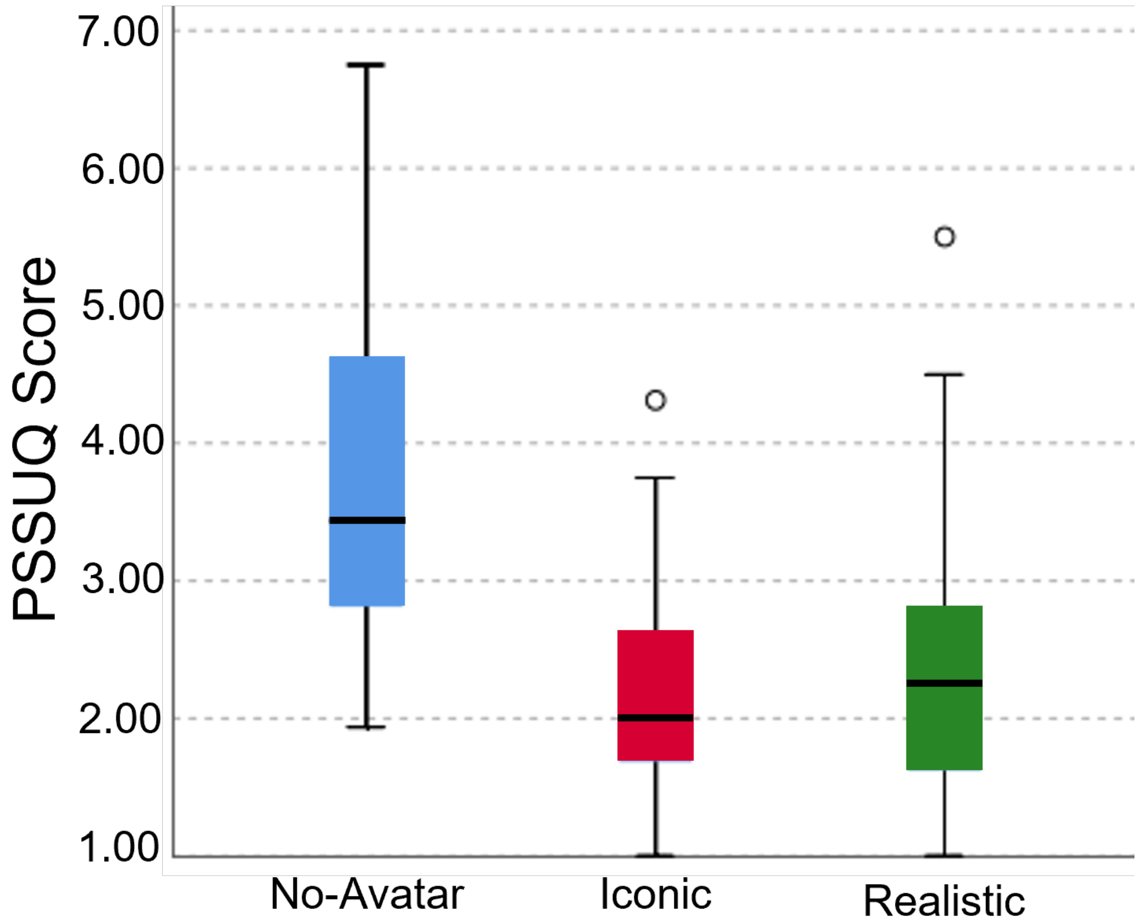


Figure 4.9: Effect of condition on perceived usability. A lower PSSUQ corresponds to a higher perceived usability. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

4.4.4 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 4.9 depicts these results.

4.5 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) [67]. 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 would be where their real hand actaully 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've had 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 density of obstacles (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. 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. Together, the results align with findings obtained by the authors of [209], who employed a similar task. Thirdly, we found that the retrieval of deeper targets resulted in poorer performance, offering support for 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 effects 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. We found that the effect of obstacle density on the performance was moderated by the effect of condition. Figure 4.6 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' actual physical (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 [224]. 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 virtual light intensity have a 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 [6]. 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 [58, 236]. 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.

4.6 Conclusion and Future Work

In this work, we investigated how avatarization affects users' performance in a near-field interaction task in augmented reality. Users were tasked with retrieving 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 avatar representation across three experimental conditions, namely, No-Avatar (no hologram overlaid), Iconic Augmented Avatar (overlaying an avatar represented as tracked joints and bones), and Realistic Augmented Avatar (overlaying an avatar represented as a tracked hand). We further examined the effects of obstacle density, obstacle size and virtual light intensity, manipulating these factors within participants. Results indicated that users in the two avatar conditions performed significantly better than users who weren't provisioned an avatar representation. Furthermore, users perceived the interaction with the system to be less usable without an avatar than when provisioned with one. Additionally, we found that the level of virtual light intensity applied affects how visible users' real hands are to them. Overall, we find that avatarizing users' end-effectors, by providing a visualization of the interacting layer is highly desirable for 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 such 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.

Chapter 5

Evaluating the Effects of the Physicalities of Interacting Components, End Effector Representations, and Interaction Techniques on Near-field Mixed Reality Interactions

The work presented in this chapter was supported in part by the US National Science Foundation (CISE IIS HCC) under Grant No. 2007435.

5.1 Motivation and Overview

The field of mixed reality (MR) has witnessed significant advancements in recent years, revolutionizing the way we interact with virtual content overlaid in the real world. Technological conglomerates like Apple, Meta, and Google are actively competing with each other to produce high-quality mixed reality HMDs that allow users to interact with virtual (augmented) content with the physical world in mind. With mixed reality, virtual components that form part of the experience have the added capability to interact with physical components in the real world. This enhanced layer of physical interactive capability makes it immensely important to study how interactions take place in this medium. A number of AR and MR applications feature interactions that take place in users' peripersonal spaces, and many of these require precise perception-action coordination. Examples of such include picking up and manipulating tools for surgical and training operations, pressing the keys on a keyboard, moving and dragging icons, menus, and items in the immersed workspace to name a few. In these kinds of applications, a small drop in accuracy can have serious consequences, making precision an aspect of high priority. Unlike pure virtual reality, near-field interactions in these sorts of mixed reality applications are primarily based on gestures made by the user's hand which serves as the primary means to engage with the digital/augmented content. Given that interactions in such mediums involve users' actual(physical) hands and require fine motor control, it is conducive to study how users' self-avatars affect these interactions given what our previous study has shown. Furthermore, it is important to study how self-avatar representations affect users' interaction performance in areas that extend just beyond their natural workspace available to their real hands. This extension of the users' natural workspace can be realized by provisioning them with self-avatars that feature translational gains and offsets. In this featurette, as the user move their actual hands, their self-avatar moves a larger distance (based on the proportion of distance from a predefined origin), allowing users to reach virtual artifacts located further away from their natural reach envelope. Investigating this in the context of perceptual calibration, a process wherein users' performance on perceptual tasks can be calibrated based on feedback, is also an area worth researching. Additionally, it remains to be seen as to what effects the physicality of the interacting components has on mixed reality interactions, especially when provisioned with an avatar. Besides, different types of interaction techniques/metaphors may be more suited for different types of tasks. In line with these broad interests, this study investigates how self-avatar representations, interaction techniques, and the physicality of interacting content collectively interact to affect near-field fine-motor perception-action coordination in mixed reality.

In this contribution, we empirically evaluated how different virtual end-effector representations affect users' performance in a near-field object manipulation-based peg-transfer task. Users were tasked with manipulating a holographic ring, transferring it from one peg to another for a number of trials while avoiding collisions with the pegs. 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 (no holograms overlaid on physical components); (2) Virtual+Physical (holograms overlaid on physical components); (3) Virtual (only virtual holograms without physical components). Users in each condition performed the peg-transfer task for a number of trials over a

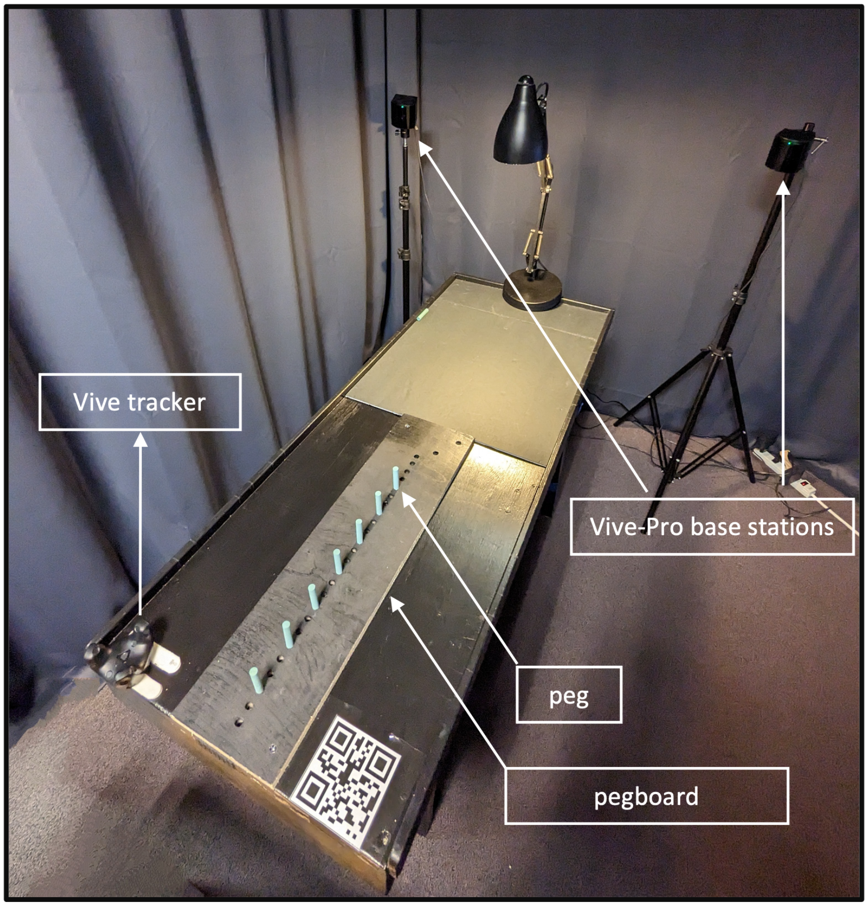


Figure 5.1: Physical Apparatus and setup used in the experiment. The black wooden pegboard base is screwed on to the table with 7 pegs slotted into the holes of the board. A vive-pro tracker, base stations, and a QR code was used for the calibration routine to register the virtual holograms in their respective positions in accordance with the physical setup.

number of phases in which the avatar representation and the interaction technique (metaphor used to grasp and release objects) were manipulated as within-subjects factors. We go on to discuss the implications of these representations and interaction techniques on users' performance, subjective perceptions, and preferences, detailing scenarios that merit using certain representations and techniques over others.

5.2 System Description

5.2.1 Physical Apparatus

For this experiment, a customized rectangular physical wooden pegboard base was built using a birch-sanded plywood board. The sanded plywood procured was uniformly smooth throughout its surfaces.

The board was carefully cut to be 15.77cm wide, 61cm long, and 1.34cm thick using a precise electrical wood-cutting saw. This board was spray painted with a matte black color uniformly throughout. An electrical 9-inch bench drill press machine equipped with a drill bit was used to drill holes of radius=95mm along a straight line running down the middle (along the depth axis) of the wooden peg board. The holes were precisely drilled such that the distance between the centers of every two adjacent holes was exactly 2cm apart. Seven identical cylindrical wooden pegs were made by precisely cutting wooden dowels using a specialized electrical variable speed scroll saw [228]. The radius of the pegs was 95mm and their heights were 6.344cm. Each peg was spray painted carefully using a coastal sage matte color, ensuring that the amount of coating did not unduly affect the thickness of the pegs. The pegs were hence designed to intricately slot into the holes of the wooden board and be tightly fastened to prevent any movement of the pegs. Given that the radius of the pegs and the holes were precisely equal, the bottom of each peg had to be sanded using sandpaper to allow the peg to slot into the holes. After sanding was completed, the pegs would neatly slot into any of the holes after which no movement of the pegs occurred even when being touched physically. This ensured that the pegs were rigidly locked in place on the peg board. The height of the pegs was chosen to be 6.34cm accounting for the 1.344cm thickness of the wooden peg board, thus making the height of each peg from the base of the board 5cm after it was slotted into the hole. After being slotted into the peg board, the height of each peg was measured (sampled) 10 times using a vernier caliper to check for consistency and accuracy in peg heights. The mean height of each peg's height as a result of these 10 measurements was calculated to be 49.82cm, 49.92, 49.96cm, 49.37cm, 49.52cm, 49.39cm, and 49.30cm respectively. The wooden peg board was mounted and secured on top of a physical wooden table (80cm tall) by drilling in screws through the board and table. The front edge of the peg board was flush with the edge of the table. With several equally spaced holes that also have small center-to-center distances, the apparatus overall, offers the potential for manipulating the distance between pegs and the distance of the first peg from the user, making for a customizable peg-placement board that can be used to study fine-motor, near-field interactions in mixed reality settings. In this study, the pegs were slotted into holes such that the first peg was 7cm away from the start of the board and the distance (center to center) between any two successive pegs was 6cm. A lamp was placed on the far end of the wooden table to facilitate better tracking and viewing of the apparatus when wearing the head-mounted display. This physical apparatus is depicted in figure 5.1, while the schematic representation of the pegboard, the pegs, and the ring is shown in figure 5.4.

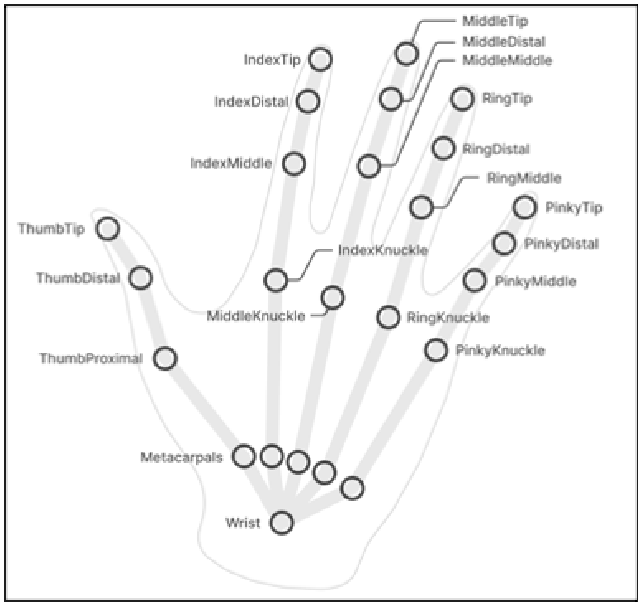


Figure 5.2: Schematic representation of the joints and phalanges tracked by the MRTK framework on the Microsoft Hololens 2.

5.2.2 Hardware and Equipment

The mixed reality simulation used for this study was built using the Unity 2020.2.2f1 game engine software and was rendered on a Microsoft Hololens 2 optical-see-through HMD using a computer equipped with an Intel i7-8700 processor, 32 GB of RAM, and an NVIDIA RTX 2080 graphics card. The Hololens 2 HMD 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), allowing to simulate more accurate spatial sounds from precise locations associated with the scene. Users were seated on a chair in front of a physical table atop which sat the wooden peg-board base and the pegs that were used in this study. The table top was painted black to facilitate better viewing experiences of the augmented holograms. During pilot testing, the simulation's frame rate was measured, ensuring that it was stable and approximately equal to the device's maximum refresh rate (60Hz).

5.2.3 Virtual Components (Holograms)

The holographic components used in this study included the ring and the pegs. The ring was modeled such that its outer diameter was 4cm, its inner diameter was 3cm, and its thickness, 0.75cm. As a result, the hole in the center of the ring was 3cm in diameter. A uniformly patterned, non-solid white texture was applied to the ring to make its contour appear salient. The virtual cylindrical pegs were modeled and scaled to match the dimensions of their real-world physical counterparts. The color and texture applied to the virtual pegs were matched to the color used on the physical pegs by using an image of the physical texture captured on a high-resolution iPhone 11 camera (240 fps). For the condition where the pegs were purely physical, a custom shader was created and applied to the virtual pegs such that they appeared invisible (transparent) yet occluded the ring appropriately. This gave the appearance that the actual physical pegs occluded parts of the ring when seen through the display. Essentially, this meant that the virtual pegs were co-located with their physical counterparts and were still present in the scene unbeknownst to the user because they were rendered transparent yet used for occlusion and collision events when the pegs were physical. All the virtual components were modeled precisely to scale using the Autodesk Maya 2019 3D modeling software.

5.2.4 Calibration and Registration Routine

A three-pronged approach was used as a step-wise calibration routine in order to establish and register the precise location of the virtual pegs on the physical pegboard. This calibration routine also helped ensure that the location of the virtual pegs on the pegboard was near-perfect and consistent regardless of the physicality of the pegs.

First, a VR simulation was run on a machine equipped with the Vive Pro Tracking system (base stations and trackers) and this was used to register the virtual apparatus on the pegboard as a first step. To accomplish this, an HTC Vive tracker was secured to the physical table and affixed firmly to a specific position (see figure 5.1) and remained in this position throughout the course of the study. The position of the Vive tracker was at a predefined distance from the peg board. 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. These positions were used in conjunction with the playspace determined by Hololens 2's inside-out tracking system to register the holographic apparatus on the peg board, completing the first step. Next, the experimenter made fine adjustments to the position and orientation of the virtual scene using a keyboard based on feedback from the participant. The precision of the translational adjustments could be as low as 1mm, and 0.1 degrees for the rotational adjustments. These adjustments were made in order to position the virtual pegs such that they manifested at a particular location on the pegboard (third hole from the edge). After final adjustments were made, a visuo-haptic verification technique was performed to ensure near-perfect co-location. In this step, 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 with the registered virtual pegs. Similarly, they would then run their index finger down each virtual peg ensuring tactile feedback as a result of the physical peg being co-located. For conditions where the pegs were completely virtual, matching physical and virtual QR codes were used 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 rendered transparent yet still remained in the exact position to account for interactions in the experiment. Furthermore, the positions 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 calibration 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 and that the lighting of the real world remained unchanged.

5.2.5 Hand Avatar Representations

This study investigated two different augmented hand avatar representations the specifics of which are described in this section. Note that an augmented hand avatar representation involves the provision of a tracked avatar (hologram) based on computations made on where the headset detects and tracks the user's actual hand using camera vision in real time. The two augmented avatar representations were compared to a baseline without any augmented avatar, making a total of three avatar representations. These representations are depicted in figure 5.3

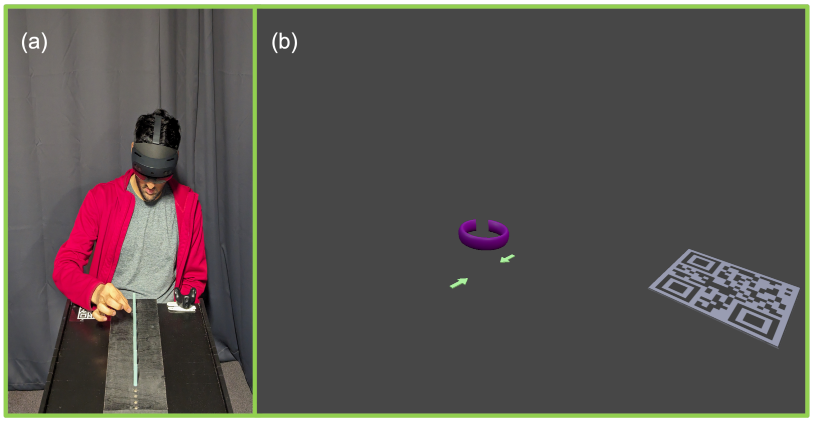
No Augmented Avatar (No-Av): In this condition, participants interact without the provision of any augmented (visualized) avatar hands. Given that the technology investigated involves optical see-through display-based mixed reality, users can see their own real (actual) hands and base their interactions on the same. A similar approach was used in the second study.

Augmented Avatar (Av): 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 skeletal representation of one's hand. The bones and joints are rendered white to make the augmented avatar salient. The augmented avatar is animated to move in real-time based on the user's own hand movements tracked by the AR system's interacting layer. 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 (see figure 5.2). A small white 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.

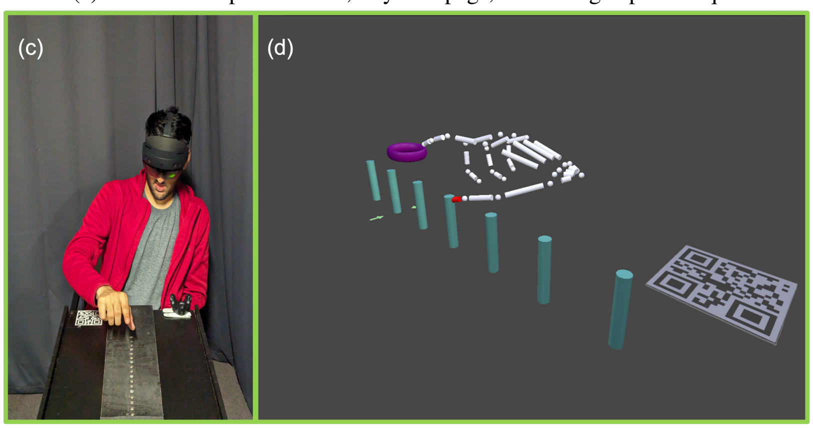
Augmented Avatar with Translational Gain (AvG): In this condition, participants interact with holograms with an augmented (visualized) hand avatar whose movements are programmed to move twice as much as the user's actual hand along the movement axis (depth axis running from the first peg to the last) in relation to a predefined origin. At the origin, the avatar is co-located with the actual hand of the user, 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 10cm 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 movement and the avatar's movement and allows for users to extend their interaction space beyond just 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 [27, 29, 104, 184]. The Avatar-gain representation's appearance is identical to the Augmented Avatar representation's appearance, except that it undertakes a positional offset in relation to the origin.

Interactions were facilitated based on computations associated with each specific avatar representation. When provisioned with an avatar, interactions with the ring and with the pegs are based on the visualized avatar. The Augmented avatar representation (Av) features an augmented visualization that visually represents the HMD's tracking of the hands, while the Avatar gain representation (AvG) visualizes the tracked hands with a gain function. Without an avatar however (No-Av), interactions between users' hands with the ring and the pegs are based on where the HMD tracks the hands to be. A recent study investigating the effects of provisioning an avatar in AR experiences on task performance used an identical approach for the No-Avatar and Avatar representations [219].

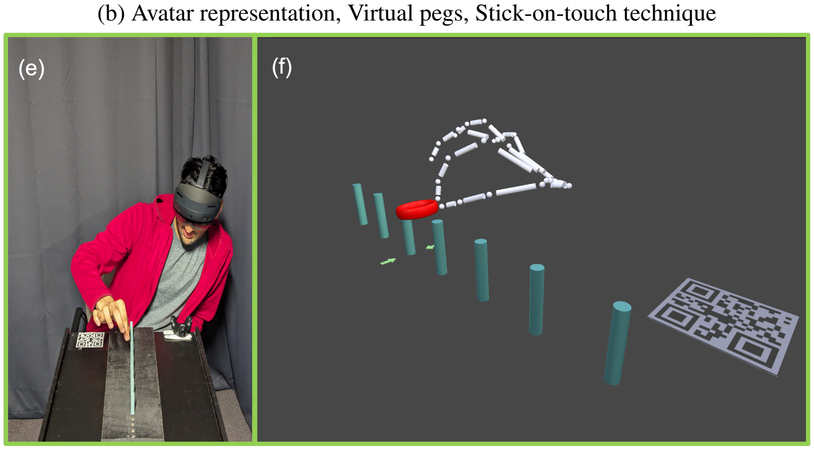
An assessment of the system was conducted to determine the positional discrepancy between the augmented avatars and the actual hand positions of users. To measure this offset, the distance between the fingertip of the user's real hand index finger and the corresponding fingertip of the avatar was measured ten times using a ruler. The measurements were taken while the user's physical hand was resting on the table. The



(a) No-Avatar representation, Physical pegs, Pinch-to-grasp technique



(b) Avatar representation, Virtual pegs, Stick-on-touch technique



(c) Avatar-Gain representation, Virtual+Physical pegs, Pinch-to-grasp technique

Figure 5.3: 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 (b) depicts the ring being centered around a physical destination peg when manipulated using the Pinch-to-grasp technique while (d) shows an avatar collision with a virtual destination peg while the ring is being manipulated using the Stick-on-touch technique. Subfigure (f) shows a ring collision with virtual+physical destination peg when using the Pinch-to-grasp technique.

average positional offset calculated from these measurements was determined to be (M=3.66mm, SD=0.69).

To evaluate the latency and frame rate in all three conditions, a system evaluation was performed using the method developed by Niehorster et al. (2017). Ten samples of latency and frame rate were measured for both simple translational and rotational hand movements in each condition. A high frame rate camera (240 fps) on an iPhone 11, mounted on a stand, was used to record videos of the user's hand and the augmented end-effector representations as seen through the viewport of the head-mounted display (HMD), with the visor removed. The user's hand was moved in a straight line for translational movements and rotated about the vertical axis for rotational movements, capturing multiple samples of each. By analyzing the recorded footage using video editing software, latency was calculated by comparing the number of frames it took for the hand movement/rotation to occur with the corresponding movement/rotation of the virtual counterpart across multiple trials.

The analysis revealed that the mean frame rate for all conditions, measured using the diagnostics profiler in the Hololens 2, was 60Hz, which matched the display's refresh rate. The mean end-to-end latency for each condition was as follows: No-avatar (Position lag = 29.16ms, Orientation lag = 28.33ms), Augmented-avatar (Position lag = 29.58ms, Orientation lag = 28.75ms), Augmented-avatar-gain (Position lag = 29.58ms, Orientation lag = 27.91ms). These values are consistent with the observed phenomenon where the hologram of the tracked hand lags behind the real hand during movement, which is supported by previous studies that have reported similar findings [201].

5.3 Experiment

5.3.1 Task

For this experiment, a peg-transfer task based on the Fundamentals of Laparoscopic Surgery (FLS) training scenario was conceptualized wherein users transferred a ring from one peg to another for a number of trials. This task serves as a basic training and evaluation tool for hand-eye coordination in laparoscopic surgical training settings [63, 203]. Users were seated on a chair in front of the table and were tasked with manipulating a holographic ring, transferring it from one peg to another 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, screen parallax, haptic feedback, and sensory mismatch on near-field fine motor perception-action coordination in virtual reality [27–29]. We drew inspiration from this task given the fine motor control and precise perception-action coordination required to perform it. The experiment was divided into six phases each of which had a specific combination of the avatar representation and interaction technique described in 5.3.2. Each of these phases commenced with three practice trials. These practice trials were designed to allow users to familiarize themselves with the mechanics associated with grasping and 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 toward that peg. These arrows together served as the indicator of the destination peg for each trial. Users would then grasp and manipulate the ring from the peg it was currently on and place it on the destination peg. When successfully placed on the destination peg, the ring turned yellow, the two arrows pointing to the peg disappeared, and a completion sound was deployed via the HMD indicating the completion of that trial. The completion of a trial 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 prior research suggests that users prefer auditory feedback rather than simply having visual feedback when grasping [36]. 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. The auditory feedback provided for these types of collision events was different from the ring collision sounds.

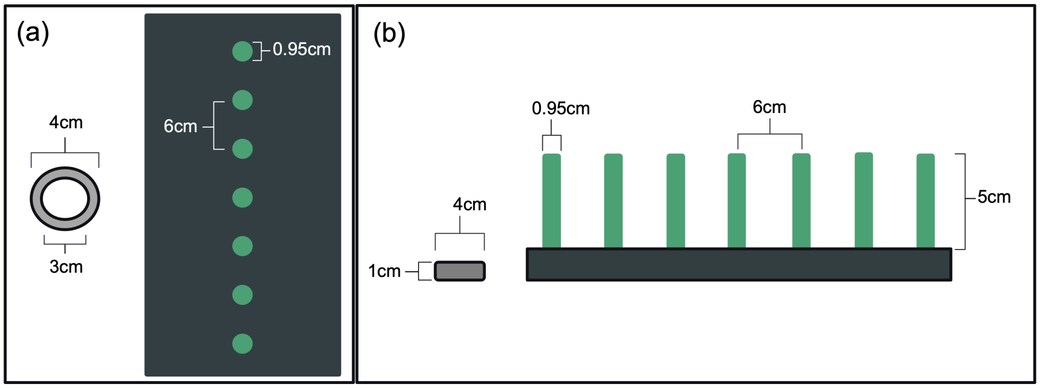


Figure 5.4: Schematic representation of the pegboard and the ring showing their dimensions (a) from a top-down perspective and (b) from a side-view perspective. Note that figures are not drawn to scale.

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 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 to ensure high degrees of fidelity in interaction design. The auditory feedback that was provided also took into account the location at which the collision event was taking place due to the fact that auditory information provided in XR affects depth perception [86, 122]. This meant that 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. Moreover, the visual and auditory feedback pertaining to any event was presented simultaneously, thus providing users with rich multi-modal feedback that was clearly indicative of the different events that transpired during a trial.

5.3.2 Study Design

To empirically evaluate how the physicality of interactive artifacts affect users' performance in a near-field, collision-avoidance, 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 5.3.1 for a number of trials over a number of phases in which the avatar representation and the interaction technique (metaphor used to grasp and release objects) were manipulated as within-subjects factors.

Two different interaction techniques categorized as Pinch-to-grab and Stick-on-touch were tested in this study. In the Pinch-to-grab interaction technique, 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. In the Stick-on-touch interaction 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 [26, 168]. For each of these interaction techniques, three different avatar representations detailed in section 5.2.5 were tested, making a total of six 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 peg-transfer 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 and it was ensured 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 every 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.

5.3.3 Measures

A number of measures gathered in this study serve as general indicators of user performance in the interaction task. In this study, the ring serves as the manipulated object, while the pegs serve as the target objects. In evaluating the performance of participants in this pick-and-place interaction task, we assess a number of variables associated with performance overall. Specifically, we focus on the efficiency, accuracy, and movement economy linked with performing the task. Efficiency deals with the time taken to manipulate the object. Accuracy, as the name suggests, pertains to how accurately users manipulate the object around the target object considering both the number of collisions between them, as well as the distance between them at the end of the trials. Both, the manipulated object, and users' end-effectors have movement economies associated with them. This movement economy pertains to the distance traveled by them during manipulation. The following portion of this section describes the variables associated with these performance-related measures in greater detail.

Number of Ring Collisions (Accuracy) - The total number of collisions associated with the ring and the pegs were measured for each trial, serving as an operational measure of accuracy. This measure was computed by incrementing a counter whenever any solid part of the ring collided with any of the pegs. Essentially, a higher number of ring collisions corresponds to worse performance.

Error Distance (Accuracy) - For each trial, the distance between the center of the ring and the center of the destination peg was measured during the release of the ring. The maximum error distance possible for any given trial was numerically equal to the inner radius (1.5cm or 0.015m) of the ring, while an error distance of 0cm corresponds to a perfectly centered ring on the destination peg. The smaller the error distance, the more accurate users are.

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.

Number of Avatar Collisions - The total number of collisions associated with users' avatar representations (task-dependent end-effectors) and the pegs were measured for each trial. Without an avatar, the tracked end-effector's collisions were measured. This measure was computed by incrementing a counter whenever any tracked joint/bone of the task-dependent effector ring collided with any of the pegs. Essentially, a higher number of avatar collisions corresponds to worse performance.

Distance traveled by the ring (Movement Economy) - In each trial, the total distance traveled by the ring when being manipulated by the user was calculated and served as an operational measure of the manipulated object's movement economy. A lower value of this variable implies that users are more economical in the way they move (manipulate) the ring.

Distance traveled by the end-effector (Movement Economy) - In each trial, the total distance traveled by the user's end-effector was calculated and served as an operational measure of the end-effector's movement economy. A lower value of this variable implies that users are more economical in the way they move their task-dependent end-effectors.

Perceived Workload - Users' perceived level of workload as a result of the simulation was measured using the NASA TLX questionnaire [74].

Perceived Usability of interaction - Users' perceived level of usability associated with interactions in this mixed reality experience was measured using the PSSUQ inventory (Post-Study System Usability Questionnaire). This is a 16-item standardized questionnaire, one that is widely used to measure usability and perceived satisfaction. Counterintuitively, a lower score corresponds to greater perceived usability [116].

5.3.4 Research Question and Hypotheses

In this study, we investigated the effects of avatar representations on a peg-transfer-based mixed reality interaction scenario with both virtual and physical components. We specifically aimed at answering the following research question: **"How does the relationship between the visualization of users' end effectors and the physicality of pegs affect interaction performance in a near-field object manipulation-based peg transfer task?"** Downstream of this major theme, we were also interested in ascertaining how interactions using such representations are affected by the interaction technique associated with grasping and releasing the manipulatable object. We operationalize performance based on the measures described in section 5.3.3. Based on this overarching research question, we developed the following hypotheses that reflect the work discussed in section 2.2, 2.5, 2.6, 2.7, and 2.9 in conjunction with our findings from the second study:

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: Users will improve their accuracy in terms of avoiding ring collisions over trials for all three avatar representations.

H4: Using the Pinch-to-grasp technique will result in higher accuracy and efficiency.

H5: Users will calibrate their efficiency over trials for all the physicality levels of the pegs.

H6: For physical pegs, accuracy will not significantly differ between the interaction techniques.

H7: Users will collide their end-effectors with the pegs less over trials for both interaction techniques.

H8: Without a gain-based avatar representation, the distance traveled by the ring will be less.

H9: Users will perceive interactions to be more usable and less workload-inducing with an avatar than without.

H10: The pinch-based interaction technique will be associated with higher usability and lower workload.

Aspects related to the technical implementation and functioning of the hardware systems used, strongly determine what effects can be expected. Conducting such a study under the assumption that the technology operates in a way that is indistinguishable from reality will likely result in the formulation of hypotheses that are different. We took into consideration the technological limitations of the systems being investigated when generating the aforementioned hypotheses.

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. When the pegs are purely physical, users are expected to perform the task simply based on the physical pegs without having direct visual information associated with the systems's approximation of the physical pegs or as how (Venkatakrishnan et al., 2023) term it, the interacting layer. Purely physical pegs are hence expected to generally result in lower accuracies. With respect to the avatar representation, prior research suggests that translational gains applied to virtual end-effectors negatively affect performance in terms of efficiency, accuracy, and the economy of motion [27, 29]. Our second study has shown that co-located avatars, which essentially represent the interaction layer, yield superior performance. Based on these previous research efforts, it is expected that the Avatar representation will produce the best performance overall and will be perceived as more usable. In terms of the interaction techniques, the pinch technique is more intuitive than the stick-on-touch given users' familiarity with the same. The physicality of the pegs is further expected to moderate the effects of the interaction technique on accuracy. When there is occlusion from the 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 experienced from occlusion. Calibration or learning occurs in tasks that involve perception-action coordination in XR [55, 220].

5.3.5 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 5.3.2). 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 of whom identified as female, 18 of whom identified as male, and the rest as non-conforming. All but one participant were right-handed, with one individual identifying themself as purely ambidextrous. All participants had normal/corrected-to-normal (20/20) vision. Overall, VR and AR experiences did not significantly differ across conditions.

5.3.6 Procedure

Upon arrival at 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, etc. 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 5.3.1), explaining the logistics involved with progressing through the six phases of the experiment. Participants then donned the Microsoft Hololens 2 OST 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 5.2.4 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 the interactions, demonstrating the gestures required to grasp and release the ring with their avatar representation and interaction technique specific to that phase. Participants then performed 3 practice trials following which they proceeded with the peg-transfer task over the 21 trials, the end of which marked the completion of that phase. After each phase, participants filled out the NASA TLX Workload questionnaire and the PSSUQ questionnaire. Users were then allowed to take a break before proceeding to the next phase. Upon completion, participants removed the HMD and filled out a post-study preference questionnaire. 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 an hour to complete the whole procedure.

5.4 Results

The number of ring collisions with pegs, time on trial, error distance, number of avatar collisions with pegs, distance traveled by the ring, and distance traveled by the end-effector were the dependent variables considered for this analysis. 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 [82]. Since the number of ring collisions with pegs, and the number of avatar collisions with pegs are count variables, performing a Poisson regression is ideal for these dependent variables. However, while performing Poisson regressions, if the dispersion parameter ϕ is greater than 1 (overdispersion), a negative binomial regression can be used [47]. For the number of ring collisions with pegs, and for the number of avatar collisions with pegs, since ϕ was found to be 1.35 and 1.89 respectively for the model with only the fixed effects, negative binomial regression was used to fix overdispersion. 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 number of ring collisions with pegs, indicating that approximately 12% of the variance in the dependent variable 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, 0.12 for error distance, 0.3 for the number of avatar collisions with pegs, 0.04 for distance traveled by the held ring, and 0.08 for distance traveled by the avatar. Following a multilevel modeling technique is ideal in this case. 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 [199]. For all the models in the analyses below, 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 (given that there were 6 phases in the experiment, each with 21 trials). Additionally, the Pinch-to-grasp and Stick-on-touch are referred to as pinch and stick respectively for simplicity. Similarly, the avatar representations are coded based on their abbreviations mentioned in section 5.2.5.

5.4.1 Accuracy

5.4.1.1 Number of Ring Collisions

A negative binomial regression was conducted to assess the effects of physicality, avatar representation, interaction technique, overall trial number, and block trial number on the number of ring collisions with pegs. This model with only the main effects (AIC = 13341, df = 10) offered a significantly better fit to the data than did the null model (AIC = 13654, df = 3), $\Delta\chi^2(7) = 326.12$, $p < 0.001$. The model explained 22% of the variance in number of ring collisions (conditional $R^2 = 0.22$, marginal $R^2 = 0.11$). There was no significant effect of physicality on the number of ring collisions with pegs. However, there was a significant effect of avatar representation on the number of collisions, χ^2 (2, N = 4536) = 124.74, $p < 0.001$, $sr^2 = 0.04$. Collisions were significantly more for the AvG representation (M = 0.34, SE = 0.069) as compared to the No-Av representation (M = 0.06, SE = 0.070), z = 6.20, $p < 0.001$, as well as the Av representation (M = -0.18, SE = 0.071), z = 11.08, $p < 0.001$. There was also a significant difference in collisions between the Av representation and No-Av representation, z = -4.96, $p < 0.001$. There was also a significant main effect of interaction technique, χ^2 (2, N = 4536) = 132.69, $p < 0.001$, $sr^2 = 0.04$. There were more collisions when using the stick technique (M = 0.30, SE = 0.066) as compared to the pinch technique (M = -0.15, SE = 0.068).

The overall trial number also significantly affected the number of collisions, χ^2 (1, N = 4536) = 52.17, $p < 0.001$, $sr^2 = 0.02$. For every one unit increase in trial number, there was a 0.4% decrease in the expected number of ring collisions with pegs. Similarly, the block trial number also had a significant effect on the number of collisions, χ^2 (1, N = 4536) = 22.36, $p < 0.001$, $sr^2 = 0.006$. For every one unit increase in block trial number, there was a 1.5% decrease in the expected number of ring collisions with pegs.

There was a significant interaction between interaction technique and physicality, χ^2 (2, N = 4536) = 15.31, $p < 0.001$, $sr^2 = 0.007$ (figure 5.5). When testing simple effects, when participants were in the physical condition, collisions were significantly more when the stick interaction technique was used (M = 1.36, SE = 0.16) as compared to when the pinch technique was used (M = 1.05, SE = 0.12), z (1512) = 4.12, $p < 0.001$. When participants were in the V+P condition, collisions were significantly more when the stick interaction

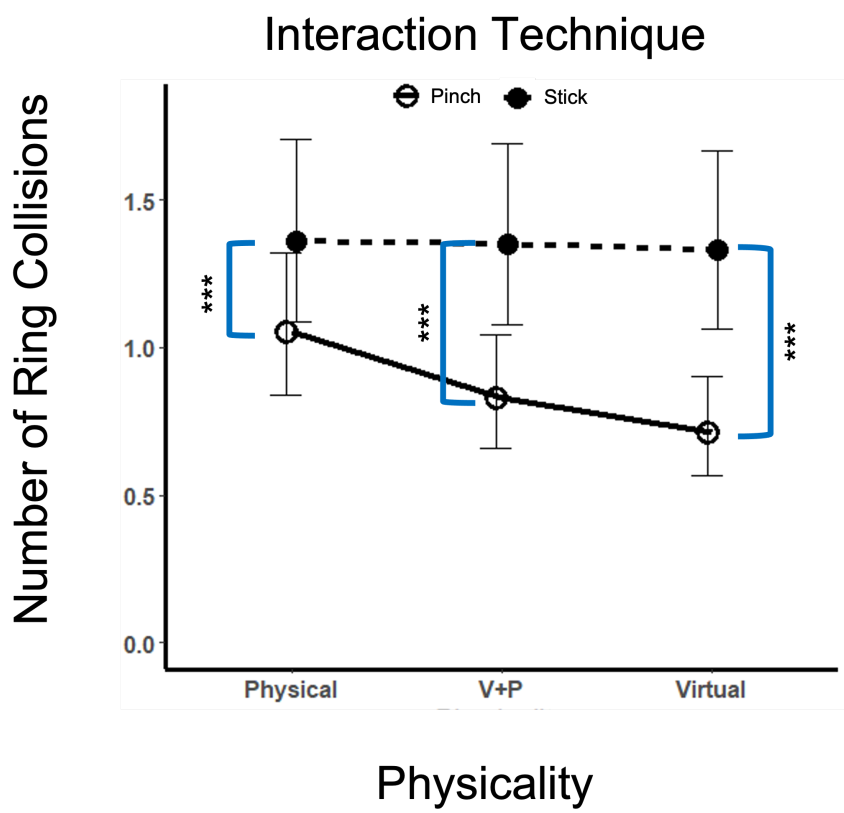


Figure 5.5: Interaction between physicality and interaction technique. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

technique was used (M = 1.35, SE = 0.16) as compared to when the pinch technique was used (M = 0.83, SE = 0.10), z (1512) = 7.62, $p < 0.001$. Similarly, when participants were in the virtual condition, collisions were significantly more when the stick interaction technique was used (M = 1.33, SE = 0.15) as compared to when the pinch technique was used (M = 0.71, SE = 0.08), z (1512) = 7.35, $p < 0.001$.

There was a significant interaction between interaction technique and avatar representation, χ^2 (2, N = 4536) = 16.81, $p < 0.001$, sr^2 = 0.008 (figure 5.6). When testing simple effects, when participants were provisioned with the Av representation, collisions were significantly more when the stick interaction technique was used (M = 1.06, SE = 0.08) as compared to when the pinch technique was used (M = 0.65, SE = 0.05), z (1512) = 6.68, $p < 0.001$. When participants were provisioned with the AvG representation, collisions were significantly more when the stick interaction technique was used (M = 1.62, SE = 0.12) as compared to when the pinch technique was used (M = 1.25, SE = 0.09), z (1512) = 4.30, $p < 0.001$. Similarly, when participants were provisioned with the No-Av representation, collisions were significantly more when the stick interaction

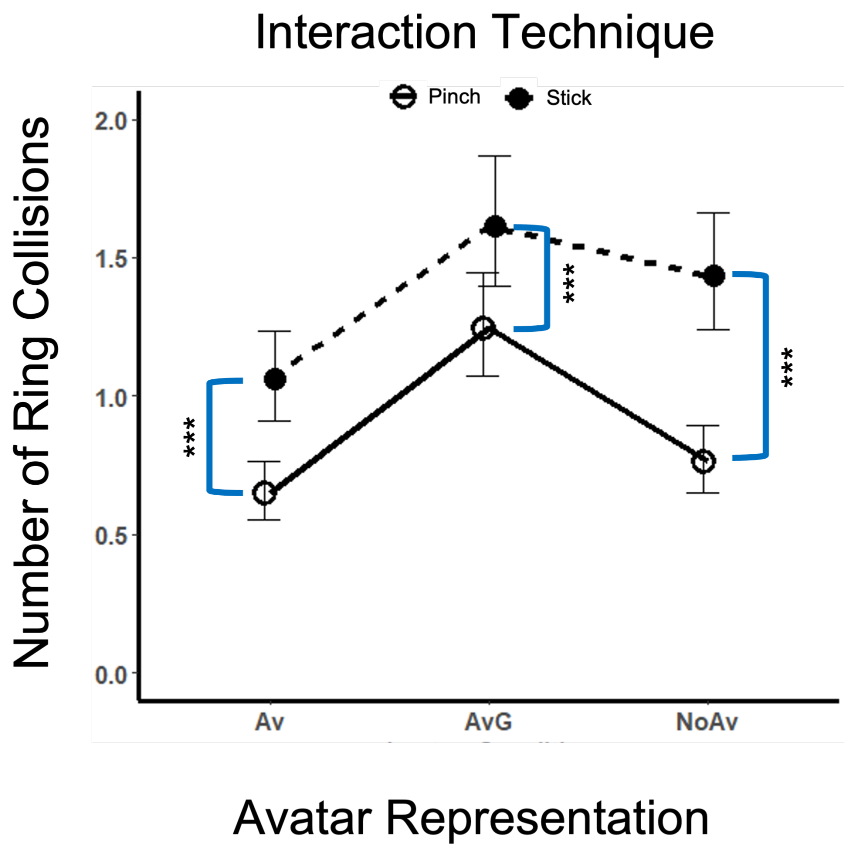


Figure 5.6: Interaction between avatar representation and interaction technique. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

technique was used (M = 1.44, SE = 0.11) as compared to when the pinch technique was used (M = 0.76, SE = 0.06), z (1512) = 8.40, $p < 0.001$.

Avatar representation was a significant moderator for the effect of overall trial number on the number of ring collisions with pegs, χ^2 (2, N = 4536) = 6.73, $p = 0.03$, sr^2 = 0.004. That is, avatar representation altered the slope (or rate of change) of the relationship between trial number and collisions. As seen in figure 5.7, a test of simple slopes revealed that for each avatar representation, the simple slope for trial number was negative and significantly different from zero. The No-Av representation (B = -0.0055, SE = 0.001, t = -5.67, $p < 0.001$) had a steeper negative slope as compared to the Av representation (B = -0.0044, SE = 0.001, t = -4.30, $p < 0.001$) and AvG representation (B = -0.0021, SE = 0.001, t = -2.34, $p = 0.02$).

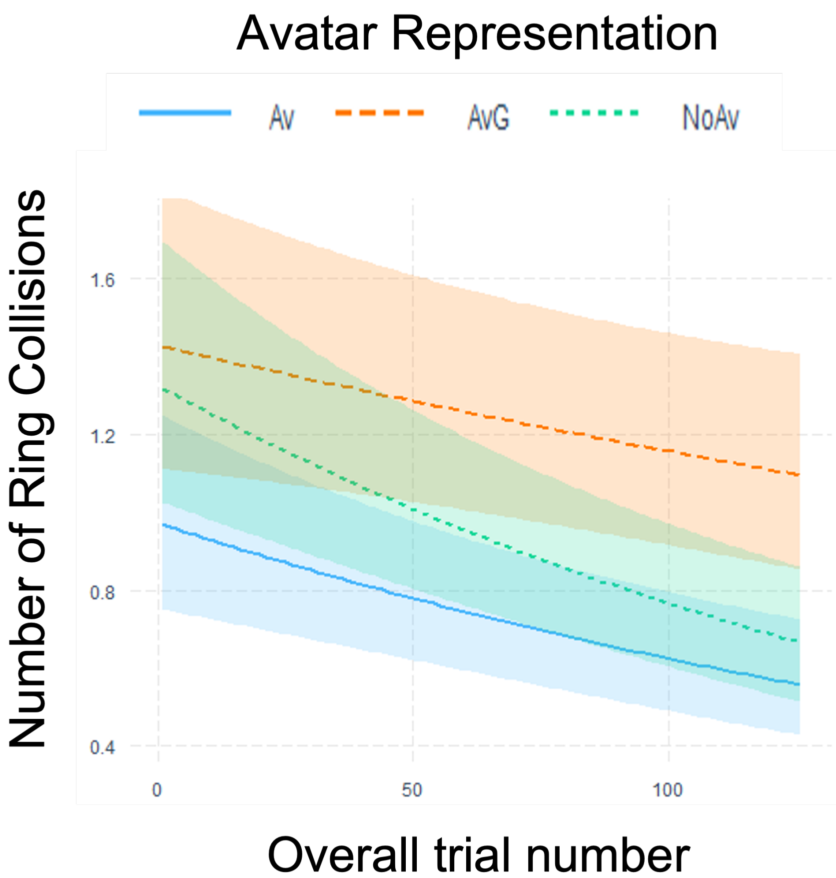


Figure 5.7: Interaction between the avatar representation and the overall trial number. Shading around lines indicate 95% confidence intervals.

5.4.1.2 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

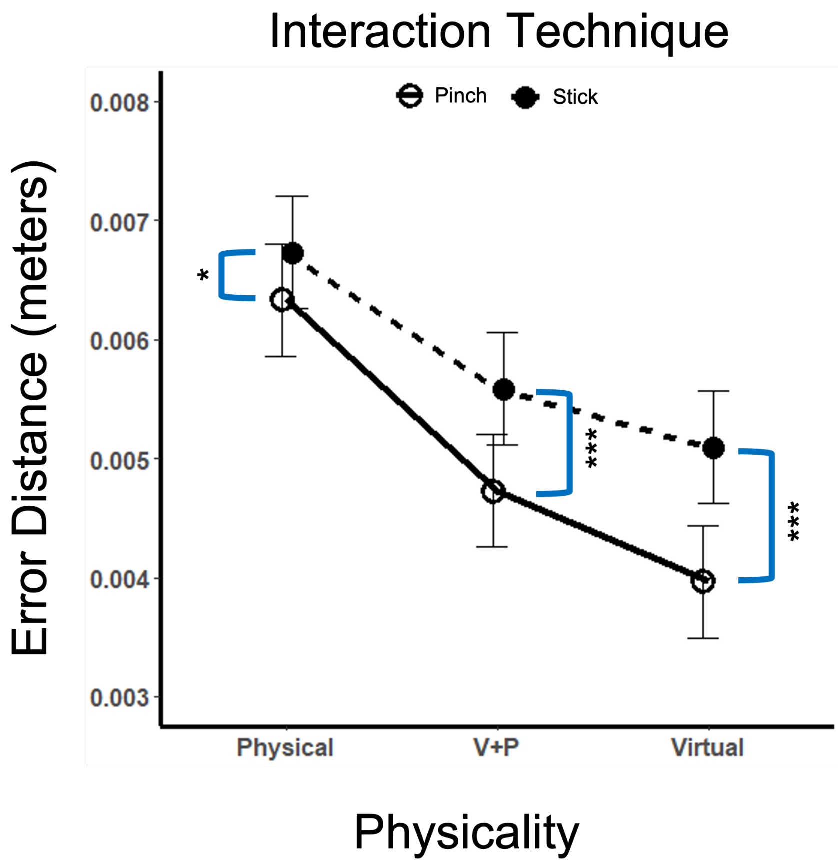


Figure 5.8: Interaction between physicality and interaction technique. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

error distance, F (2, 4495) = 16.11, $p < 0.001$, sr^2 = 0.006. Error distance was significantly more for the AvG representation (M = 0.0058, SE = 0.0001) as compared to the No-Av representation (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 representation and 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 technique (M = 0.0058, SE = 0.0001) as compared to the pinch technique (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 0.000003 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 0.00004 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 5.8). 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 was a significant moderator for 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 5.9, a test of simple slopes revealed that the simple slope for trial number was negative and significantly different from zero only for the virtual condition (B = -0.0000098, SE = 0.000002, t = -4.60, $p < 0.001$), while the slopes were not significantly different from zero for the V+P condition and the physical condition.

5.4.1.3 Number of Avatar Collisions

A negative binomial regression was conducted to assess the effects of physicality, avatar representation, interaction technique, overall trial number, and block trial number on the number of avatar collisions. This model with only the main effects (AIC = 8453.2, df = 10) offered a significantly better fit to the data than did the null model (AIC = 9128.9, df = 3), $\Delta\chi^2(7)$ = 689.72, $p < 0.001$. The model explained 61% of the variance in number of avatar collisions (conditional R^2 = 0.61, marginal R^2 = 0.45).

There was no significant effect of physicality on the number of avatar collisions. However, there was a significant effect of avatar representation on the number of collisions, χ^2 (2, N = 4536) = 27.51, $p < 0.001$, sr^2 = 0.009. Collisions were significantly different for the AvG representation (M = -1.01, SE = 0.17) as compared to the No-Av representation (M = -0.74, SE = 0.17), z = -2.41, $p = 0.04$, as well as the Av representation (M = -1.33, SE = 0.17), z = -2.85, $p = 0.01$. There were also significantly fewer collisions for the Av representation as compared to the No-Av representation, z = -5.25, $p < 0.001$. There was also a significant main effect of interaction technique, χ^2 (1, N = 4536) = 701.88, $p < 0.001$, sr^2 = 0.29. There were fewer collisions when using the stick technique (M = -2.38, SE = 0.17) as compared to the pinch technique (M = 0.32, SE = 0.16).

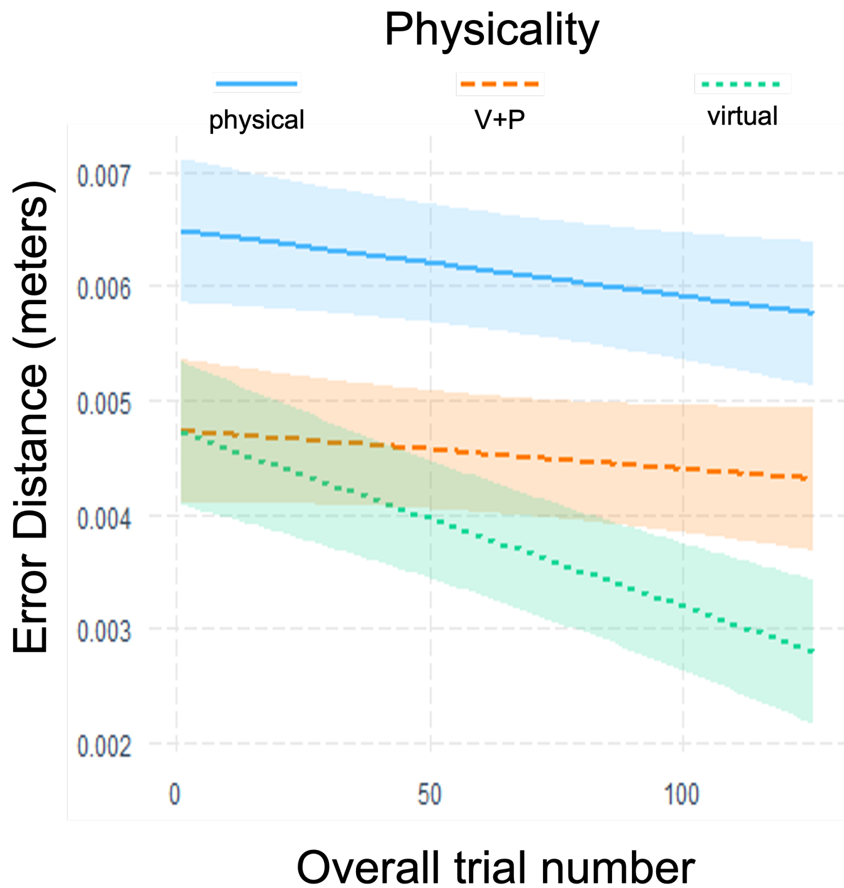


Figure 5.9: Effect of trial number on error distance, moderated by physicality. Shading around each line indicates 95% confidence intervals.

The overall trial number also significantly affected the number of collisions, χ^2 (1, N = 4536) = 95.12, $p < 0.001$, sr^2 = 0.03. For every one unit increase in trial number, there was a 1.3% decrease in the expected number of ring collisions with pegs. Similarly, the block trial number also had a significant effect on the number of collisions, χ^2 (1, N = 4536) = 12.55, $p < 0.001$, sr^2 = 0.001. For every one unit increase in block trial number, there was a 2.7% decrease in the expected number of ring collisions with pegs.

There was a significant interaction between interaction technique and physicality, χ^2 (2, N = 4536) = 18.56, $p < 0.001$, sr^2 = 0.001 (figure 5.10). When testing simple effects, when participants were in the physical condition, collisions were significantly more when the pinch interaction technique was used (M = 2.42, SE = 0.68) as compared to when the stick technique was used (M = 0.11, SE = 0.033), z (1512) = -18.27, $p < 0.001$. When participants were in the virtual+physical condition, collisions were significantly more when

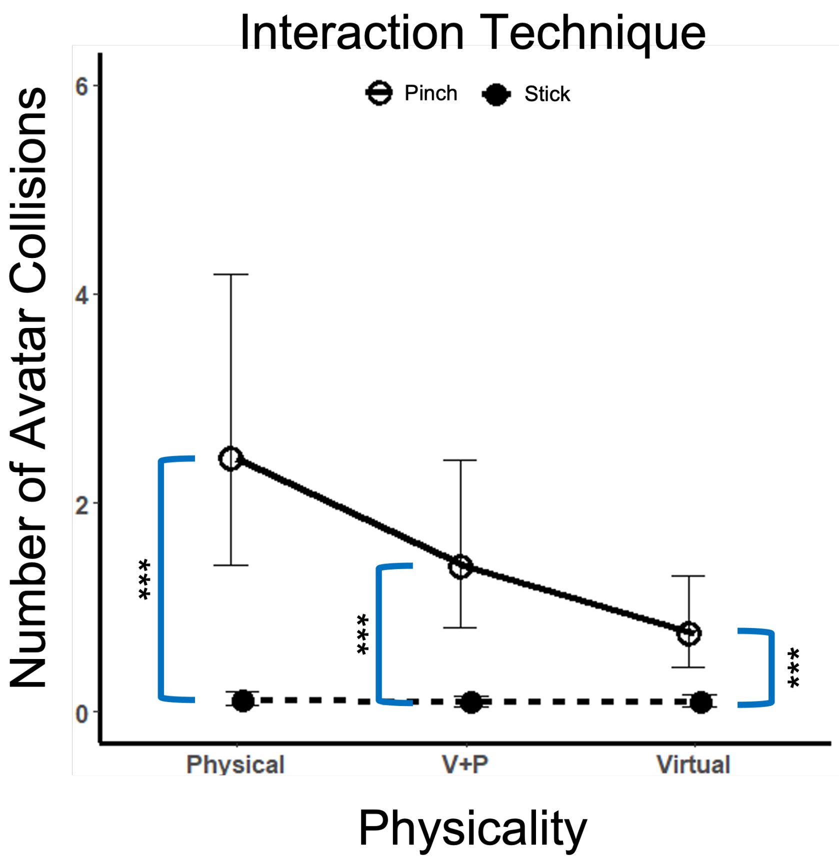


Figure 5.10: Interaction between physicality and interaction technique. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

the pinch interaction technique was used (M = 1.39, SE = 0.39) as compared to when the stick technique was used (M = 0.083, SE = 0.025), z (1512) = -15.03, $p < 0.001$. Similarly, when participants were in the virtual condition, collisions were significantly more when the pinch interaction technique was used (M = 0.74, SE = 0.21) as compared to when the stick technique was used (M = 0.093, SE = 0.03), z (1512) = -12.98, $p < 0.001$.

Interaction technique was a significant moderator for the effect of overall trial number on the number of avatar collisions, χ^2 (1, N = 4536) = 17.06, $p < 0.001$, $sr^2 = 0.01$. That is, the interaction technique altered the slope (or rate of change) of the relationship between trial number and collisions. As seen in figure 5.11, a test of simple slopes revealed that for the pinch condition, the simple slope for trial number was negative and significantly different from zero (B = -0.02, SE = 0.0025, t = -9.09, $p < 0.001$), while that for the stick condition was not significantly different from zero.

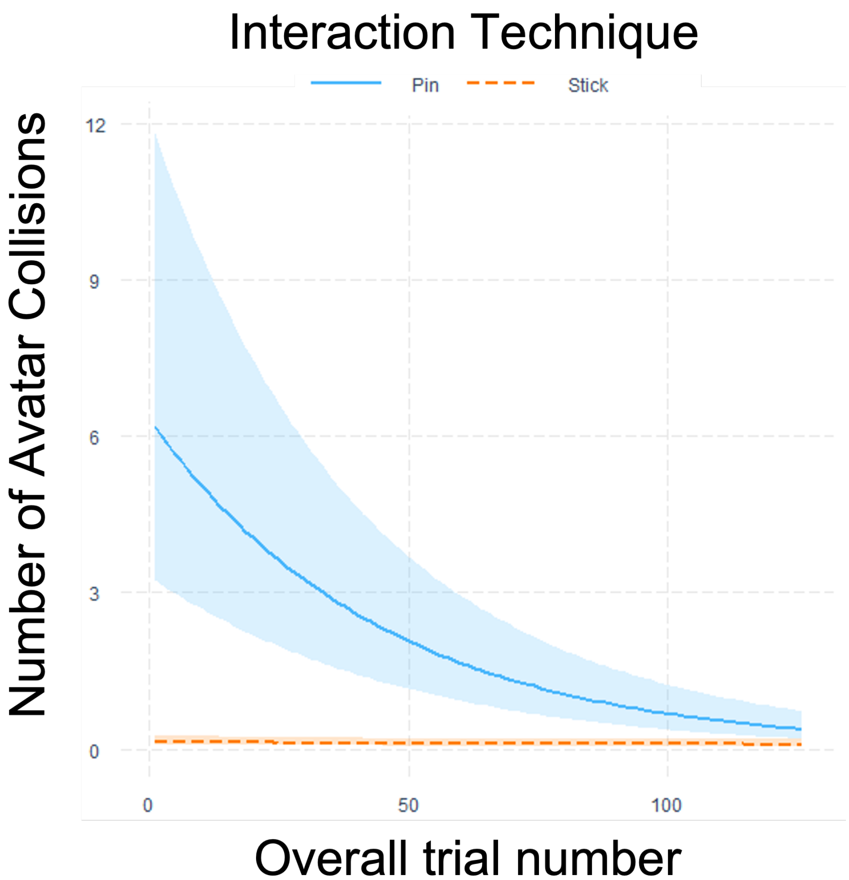


Figure 5.11: Effect of trial number on the number of avatar collisions, moderated by interaction technique. Shading around each line indicates 95% confidence intervals.

5.4.2 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

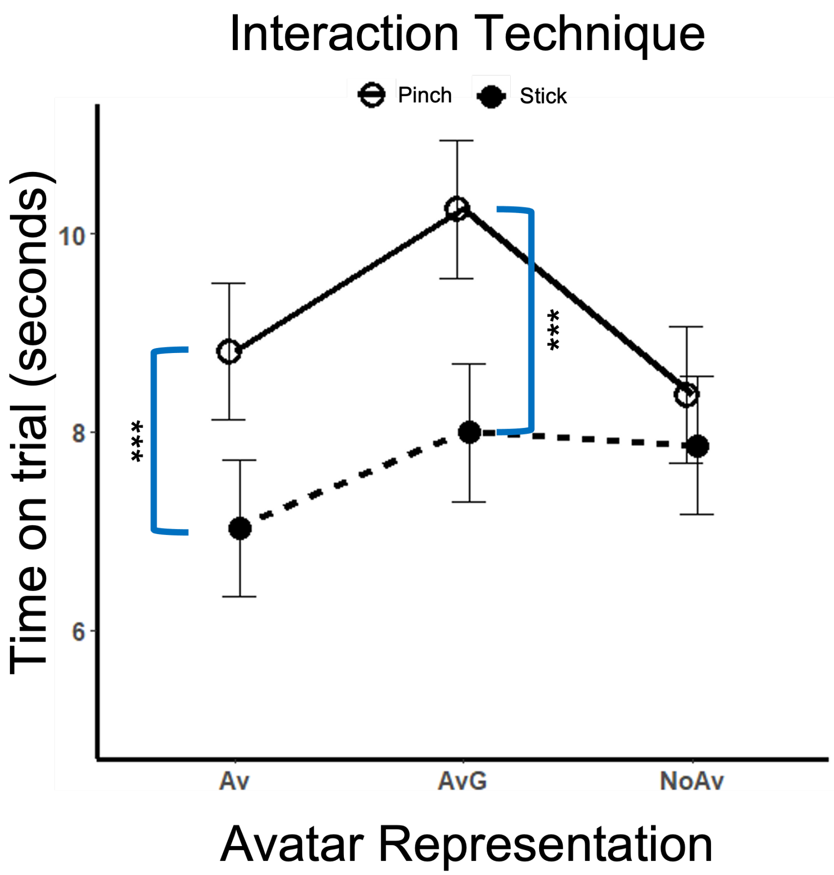


Figure 5.12: Interaction between avatar representation and interaction technique. Error bars indicate 95% confidence intervals. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

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, $p < 0.001$, $sr^2 sr^2 = 0.02$. 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 5.12). 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

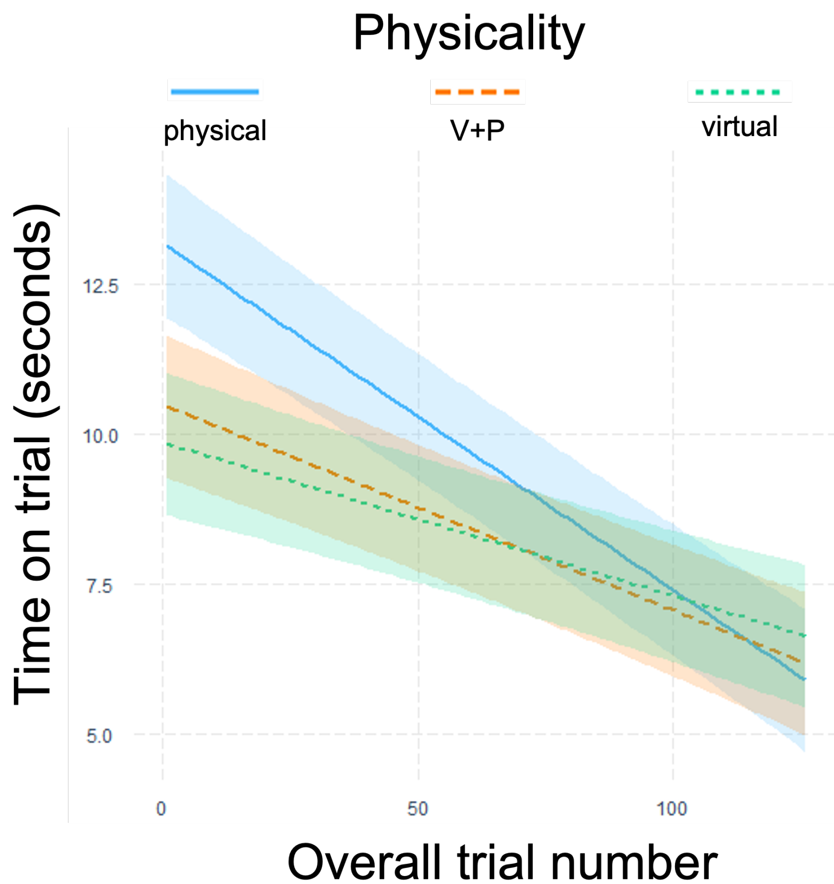


Figure 5.13: Effect of trial number on time on trial, moderated by physicality. Shading around each line indicates 95% confidence intervals.

= 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 5.13, a test of simple slopes revealed that for each physicality, the simple slope for trial number was negative and significantly different from zero. The 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, $p < 0.001$).

5.4.3 Movement Economy

5.4.3.1 Distance traveled by the ring

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 distance traveled by the ring. This model with only the main effects (AIC = 2545.34, df = 10) offered a significantly better fit to the data than did the null model (AIC = 2670.61, df = 3), $\Delta\chi^2(7) = 139.27$, $p < 0.001$. The model explained 6.6% of the variance in distance traveled by the ring (conditional $R^2 = 0.066$, marginal $R^2 = 0.031$).

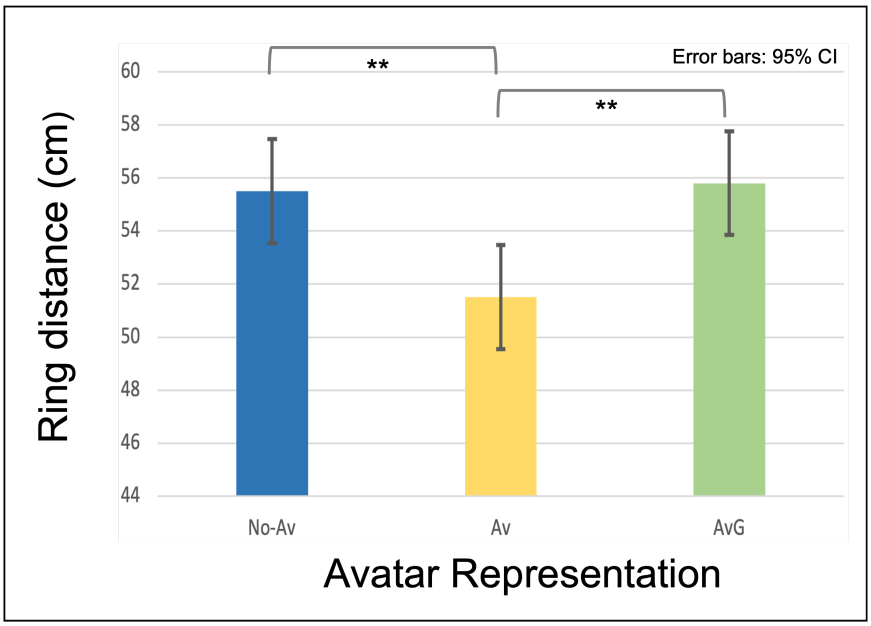


Figure 5.14: Effect of avatar representation on the movement economy associated with the ring (distance traveled by the ring). Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

There was a no significant effect of physicality on distance traveled by the ring. However, there was a significant effect of avatar representation on distance traveled by the ring, F (2, 4495) = 8.76, $p < 0.001$, $sr^2 = 0.004$. The distance was significantly lesser for the Av representation (M = 0.515, SE = 0.01) as compared to the No-Av representation (M = 0.555, SE = 0.01), t = -3.50, $p = 0.001$, as well as the AvG representation (M = 0.558, SE = 0.01), t = -3.74, $p = 0.001$. There was no significant difference between the AvG representation and the No-Av representation in terms of distance traveled by held ring. This result is depicted in figure 5.14.

There was also a significant main effect of interaction technique, F (1, 4495) = 18.99, $p < 0.001$, sr^2 = 0.004. Distance traveled by the ring was greater when using the stick technique (M = 0.56, SE = 0.01) as compared to the pinch technique (M = 0.52, SE = 0.01).

The overall trial number also significantly affected the distance traveled by the held ring, F (1, 4495) = 91.65, $p < 0.001$, sr^2 = 0.02. As the trial number increased by 1 standard deviation (SD) units, the distance traveled by held ring decreased by 0.001 SD units. Similarly, the block trial number also had a significant effect on the distance traveled by the ring, F (1, 4495) = 11.23, $p < 0.001$, sr^2 = 0.002. As the block trial number increased by 1 SD unit, the distance traveled by the ring decreased by 0.003 SD units.

5.4.3.2 Distance traveled by the end-effector

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 distance traveled by the end-effector. This model with only the main effects (AIC = 8058.83, df = 10) offered a significantly better fit to the data than did the null model (AIC = 8364.34, df = 3), $\Delta\chi^2(7)$ = 319.51, $p < 0.001$. The model explained 14.3% of the variance in distance traveled by the end-effector (conditional R^2 = 0.143, marginal R^2 = 0.080). There was a significant effect of physicality on distance traveled by the end-effector, F (2, 33) = 5.08, $p = 0.01$, sr^2 = 0.02. Distance traveled by the end-effector was significantly more in the physical condition (M = 0.717, SE = 0.05) as compared to the virtual condition (M = 0.502, SE = 0.05), t = 3.19, $p = 0.01$, but was not different from the V+P condition (M = 0.61, SE = 0.05). Distance traveled by the end-effector was also not different between the virtual and V+P condition. There was a significant effect of avatar representation on the distance traveled by the end-effector, F (2, 4495) = 23.79, $p < 0.001$, sr^2 = 0.009. The distance was significantly lesser for the Av representation (M = 0.54, SE = 0.03) as compared to the No-Av representation (M = 0.68, SE = 0.03), t = -6.90, $p < 0.001$, as well as the AvG representation (M = 0.61, SE = 0.03), t = -3.33, $p = 0.003$. The distance was also significantly different for the AvG representation as compared to the NoAv representation, t = -3.56, $p = 0.001$. There was also a significant main effect of interaction technique, F (1, 4495) = 32.05, $p < 0.001$, sr^2 = 0.006. Distance traveled by the end-effector was greater when using the pinch technique (M = 0.66, SE = 0.03) as compared to the stick technique (M = 0.56, SE = 0.03). The overall trial number also significantly affected the distance traveled by the end-effector, F (1, 4495) = 172.20, $p < 0.001$, sr^2 = 0.03. As the trial number increased by 1 standard deviation (SD) units, the distance traveled by the end-effector decreased by 0.003 SD units. Similarly, the block trial number also had a significant effect on the distance traveled by the end-effector, F (1, 4495) = 69.11, $p < 0.001$, sr^2 = 0.01. As the block trial number increased by 1 SD unit, the distance traveled by the end-effector decreased by 0.012 SD units.

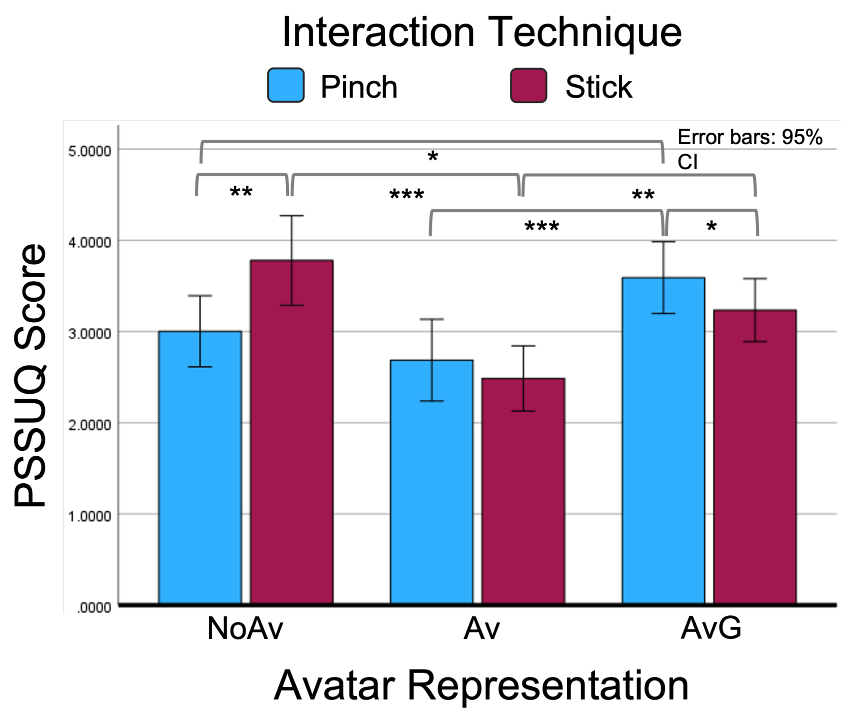


Figure 5.15: Interaction effect of avatar representation and interaction technique on perceived usability. A lower PSSUQ score corresponds to a higher perceived usability associated with interactions. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

5.4.4 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. We found a significant main effect of the avatar representation, F(2, 68) = 16.105, $p < 0.001$, $\eta_p{}^2$= 0.321. The mean PSSUQ score was significantly lower when users were provisioned with the Av representation (M = 2.58, SE = 0.169) as compared to the No-Avatar representation (M = 3.39, SE = 0.172), $p < 0.001$, and Avatar-Gain 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 the avatar representation and the interaction technique was found, F(2, 68) = 8.579, $p < .001$, $\eta_p{}^2$= 0.201. As seen in figure 5.15, without an avatar (when provisioned with the No-Av representation), the mean PSSUQ score was significantly higher when using the stick technique (M =

3.77, SE = 0.242) as compared to the pinch technique (M = 3.3002, SE = 0.192), $p < 0.01$. When provisioned with the Avatar-Gain representation, the mean PSSUQ score was significantly higher when using the pinch technique (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.

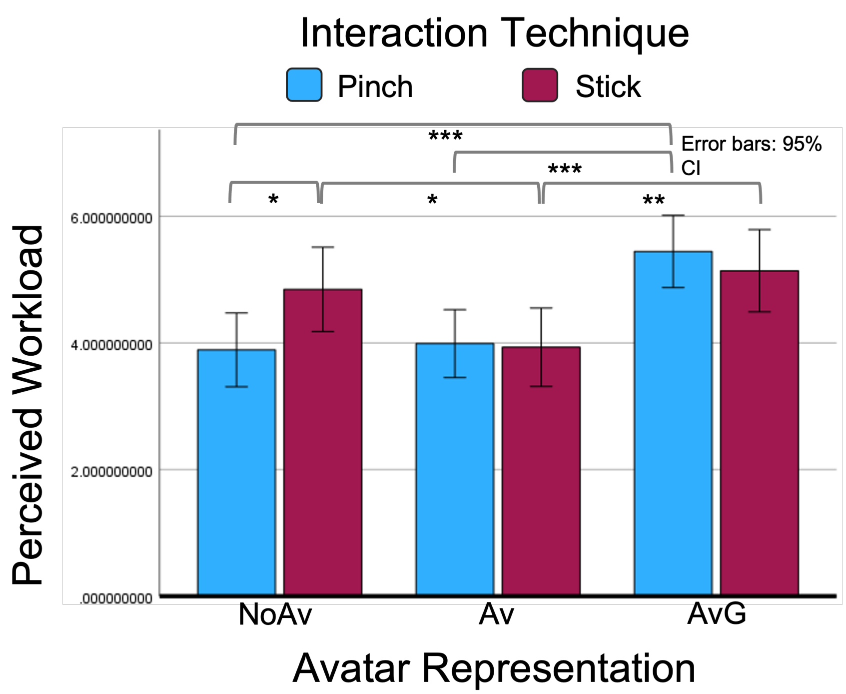


Figure 5.16: Interaction effect of avatar representation and interaction technique on perceived workload. Significance levels: '*' denotes p<0.05, '**' denotes p<0.01, and '***' denotes p<0.001.

5.4.5 Perceived Workload

A Two-way repeated measures ANOVA was conducted to determine the effects of the avatar representation and interaction technique on users' perceived workloads. We found a significant main effect of the avatar representation, F(2, 68) = 19.163, $p < 0.001$, $\eta_p{}^2$= 0.36. The mean overall perceived workload was significantly higher when users were provisioned the Avatar-Gain representation (M=5.291, SE = 0.264) as compared to the No-Avatar representation (M = 4.368, SE = 0.231), $p < 0.001$, and the Av representation (M = 3.960, SE = 0.241), $p < 0.001$. No other significant differences were found between the representations. There was no significant main effect of the interaction technique F(1, 34) = 0.845, $p = 0.365$, $\eta_p{}^2$= 0.024. However, a significant interaction effect between the avatar representation and the interaction technique was found, F(2, 68) = 4.425, $p = 0.016$, $\eta_p{}^2 = 0.115$. As seen in figure 5.16, without an avatar (when provisioned with the No-Av representation), the mean workload was significantly higher when using the stick technique (M = 4.846, SE = 0.327) as compared to the pinch technique (M = 3.890, SE = 0.296), $p = 0.024$. However, no other significant differences were found between the interaction techniques for the other avatar representations. Inspecting this interaction effect across the avatar representations, it was found that when using the pinch technique, the mean overall workload was significantly higher when users were provisioned with the Avatar-Gain representation (M = 5.444, SE = 0.280) as compared to the No-Avatar representation (M = 3.890, SE = 0.286), $p < 0.001$, and the Av representation (M = 3.989, SE = 0.264), $p < 0.001$. No other differences between avatar representations were found when using the pinch technique. When using the stick technique however, the mean overall workload was significantly lower when users were provisioned with the Avatar representation(M = 3.931, SE = 0.305) as compared to the No-Avatar representation(M = 4.846, SE = 0.327), $p < 0.05$, as well as the Avatar-Gain representation (M = 5.139, SE = 0.320), $p = 0.003$. No other differences between the avatar representations were observed when using the stick technique.

5.5 Discussion

5.5.1 Accuracy

In terms of how accurately users were able to center and place the virtual ring on the pegs, we found a main effect of the physicality of the pegs such that purely physical pegs were associated with the highest error distance in comparison to pegs that had some virtuality (V or V+P) associated with them, 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 the peg. Falling back to the idea of control laws, as explained in the previous study, it can be argued that virtual pegs make the visual information of relevance to these control laws (registered positions of the pegs) more salient, thereby allowing for increased accuracy. Simply stated, the visual information required to perform this task accurately is continuously provided when the pegs are virtual (or virtual+physical), allowing for better online control and perception-action coordination [236]. This idea aligns directly with recent research suggesting that visualizing the task-dependent information (the interacting layer corresponding to the registered pegs in this case) improves near-field interaction performance [219]. It is also important to factor in the tracking performance associated with the inside-out camera tracking system of the Hololens 2 when considering this result. Accuracy could have also been affected by mild tracking errors after registration of the pegs, especially when they were physical. Though the exact same calibration routine was employed for when the pegs were virtual, visualization of the information of relevance (the registered pegs) meant that this would be less problematic given that virtual pegs are nothing but a representation of the interacting layer. We also found a main effect of the avatar representation, in that the avatar with a translational gain (Avatar-Gain) was found to be the most inaccurate in terms of centering the ring on the pegs while avoiding collisions. It is possible that this result was observed because the gain function while potentializing the extension of one's reach envelope 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 research suggesting that adding such gains to the virtual end effector affects task performance negatively in similar scenarios [29]. Moreover, for the Avatar-Gain representation, users had to ensure that their real hand was in the field of view of the headset, ensuring that it was correctly tracked for the avatar to be rendered. This may have resulted in the users focusing on their own actual hands while performing the task with the virtually offset hand, further contributing to the aforementioned visuo-proprioceptive mismatch. Furthermore, this requirement of users having to focus the head-mounted display to track their hands in the real world may have limited the precision with which users could have performed the task with the off-set virtual hand. Eye-hand coordination, also referred to as hand-eye coordination, involves the synchronized motor control of eye and hand movements [93]. It encompasses the perception of visual information facilitated by reaching and grasping actions, while simultaneously utilizing proprioception of the hands to guide the eyes. This phenomenon represents a form of multisensory perception-action, where different types of sensory information are collectively used to guide perception and action to accomplish goals. The degree of eye-hand coordination may hence have been lesser due to this. Additionally, we found a significant effect of the interaction technique on the error distance, showing that the Stick-on-touch resulted in higher error distances than the Pinch-to-grasp technique. The Pinch-to-grasp required users to, as the name suggests, pinch the ring with their thumb and index fingertip at any one given point on its boundary. This hence resulted in users grasping the object metaphorically rather than isomorphically. The Stick-on-touch technique, however, necessitated users to manipulate the ring predominantly using their index fingertip, utilizing their thumbs only when trying to release the grasped object. These results with respect to the differences in accuracy can be explained based on the fundamentals of human hand morphology and how digits work in conjunction and isolation to perform precise fine-motor tasks. A plethora of such research suggests that precision with respect to grasping and manipulating objects is achieved using the forces of opposition (rather than isolation) that are provided by the thumb in humans and chimpanzees [34, 119, 120, 139]. Without opposition, the Stick-on-touch 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 generally less accurate in terms of centering the ring around the pegs while avoiding collisions of the ring and the pegs when grasped. The Pinch-to-grasp technique on the other hand, hence elicits the forces of opposition, allowing users to precisely grasp and manipulate the ring, thus facilitating higher levels of accuracy. Interestingly, we found an interaction effect between the physicality and interaction technique showing that the magnitude of differences between the Stick-on-touch and Pinch-to-grasp 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 5.8. 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, thus partially supporting hypothesis H6. If this was indeed the reason as to why the interaction effect was observed, both the virtual+physical and the physical conditions should not have significantly differed in terms of the magnitude of differences between the two interaction techniques. Given that we did not observe this effect, 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 increases in the physicality of the pegs were interacting to produce the observed results. Similar overall trends were observed in terms of the number of ring collisions with the pegs, another operational measure of accuracy. Overall, it appears that a concordance between the virtuality of the interacting components (manipulatable and target objects), provisioning visually co-located self-avatars depicting the interacting layer, and facilitating pinch-based interactions are desirable characteristics for increases in accuracy for such fine-motor tasks in mixed reality settings.

We observed a couple of interesting learning effects with respect to calibration of accuracy over trials. With regards to being accurate in terms of avoiding collisions of the rings with the pegs, users generally improved their performance over trials, confirming hypothesis H3. However, this learning effect was more pronounced for the Avatar and No-Avatar representations than when compared to the Avatar-Gain representation as evinced in figure 5.7. This is indicative of a diminished ability to calibrate performance when provisioned with the Avatar-Gain representation, thus partially supporting hypothesis H3. It is worth noting, however, that users were provisioned with each avatar representation twice, over two distinct phases, each having a different interacting technique. Spread out over 6 phases, this experiment required users to continually switch avatar representations after each phase. Besides, each of these phases comprised 21 trials, which could have been insufficient to see an effect of calibration for the Avatar-Gain representation. Given that the No-Avatar and Avatar representations showed this calibration effect to a larger degree under the same constraints (few trials and multiple phases), it appears that the Avatar-Gain representation requires a larger number of trials for calibration. It may hence be the case that continually and consistently using the Avatar-Gain representation without having to switch avatar representations between phases would likely yield in a desired effect of calibration over trials. The obtained findings go against results obtained in a study suggesting that adaptation to hand avatars is fairly fast [95]. Their work however investigates adaptation and calibration in an immersive virtual environment rather than a mixed or augmented reality setting. Furthermore, the avatars or hand representations that were tested in their study did not involve positional offsets or translational gains in their manifestation.

We also observed that users significantly improved their accuracy in terms of their resultant error distances over trials. However, this effect of calibration of accuracy was significantly more pronounced for the virtual condition as compared to the virtual+physical, and the physical condition as shown in 5.9. 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 physical 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 hence somewhat unreasonable to expect users to calibrate their accuracy when the explicit visual information of relevance is never there to begin with, especially given that there are and will always be discrepancies in tracking performance with any AR/MR system [67]. Besides, the feedback that users could attune to in the physical condition is relatively impoverished given that the interaction layer associated with the pegs is not visible. 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 physical condition but didn't improve at a rate similar to the virtual condition. It is likely the case that seeing virtual pegs overlaid on top of 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 errors associated with contemporary OST head-mounted displays. These are noteworthy findings given the recent growth and interest in facilitating mixed reality interactions. Moreover, 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. It is also important to consider that the type of end-effector representation provided strongly influences users' efficacy in calibrating their accuracy over time.

On the subject of the number of end effector collisions with the pegs, we observed some interesting findings. While the physicality of the pegs did not affect how frequently users collided with the pegs, the avatar representation that users were provided with impacted their collision outcomes. Users' end-effectors collided with the pegs more when they were provisioned with the Avatar-Gain representation than with the Avatar representation. Additionally, it was observed that without an avatar, users' tracked end-effectors collided with the pegs more than when provisioned with an avatar regardless of gain. The superiority of the Avatar representation can be attributed to the visual information that users have of their end-effectors along with the higher degree of precision that this representation offers. With the Avatar-Gain representation, a small movement of users' actual hands corresponds to a larger movement of the virtual end effector at farther distances, making it sensitive to slight movements and limiting the precision it offers. In the No-Avatar representation, users did not obtain visual feedback when their tracked end-effectors collided with the pegs but rather only received auditory feedback. This lack of visual feedback may help explain the inferiority observed for this representation in terms of the number of collisions of the end-effector with the pegs. Besides, occasional inaccuracies of hand tracking also led to users without an avatar getting auditory feedback of collisions (without corresponding visual feedback) even when their actual hands did not touch the pegs.

The interaction technique also significantly affected users' ability to avoid collisions between their end effectors and the pegs. It was found that users' end effectors collided with the pegs more often when using the Pinch-to-grasp than when using the Stick-on-touch technique. One possible reason for this observation hinges on the implementational specifics associated with the two techniques. When using the former technique, users were required to establish two contact points, their index finger and thumb tips, to grasp the ring. With the latter technique, however, only one contact point (the index fingertip) was needed to grasp the ring. It is hence possible that the Pinch-to-grasp technique is associated with a higher potential for collisions of the end effector and the target objects simply by virtue of the number of contact points in the vicinity of the pegs. An interesting learning effect was observed suggesting that even though users collided their hands more when using the pinch technique, they were able to calibrate their performance over the trials of the experiment (figure 5.11, suggesting that this technique is amenable to improvements in accuracy (supporting H7). Taken together with the other measures of accuracy indicating positive outcomes, these results suggest that a pinch-based interaction can be favorable even if it causes more errors (like avatar collisions with the pegs) due to its ability to facilitate calibration towards the avoidance of such kinds of flaws in performance.

5.5.2 Efficiency

The statistical analyses with respect to task efficiency suggested that users took significantly more time to transfer the ring from one peg to the other when provisioned with the Avatar-gain representation as compared to the other representations tested in this study. This is understandable given that correctly manipulating the ring, be it pinch or stick, 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 [29], directly aligning with our findings. The results observed on efficiency and accuracy (section 5.5.1) 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 were quicker or 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 H4 in terms of efficiency. This is interesting given the fact that users were significantly more accurate in performing the task of carefully placing the ring on the pegs while avoiding collisions. Another factor that may have contributed to this result of an increased time associated with the Pinch-to-grasp technique may revolve around the relative difficulty associated with grasping the ring after each trial. With the Stick-on-touch 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 the 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 in performing the manipulation task. It was found that when provisioned with an avatar (either the Avatar or Avatar-Gain representations), users took significantly more time to complete the trials when using the Pinch-to-grasp interaction technique. However, without an avatar, there was no difference between the interaction techniques used to manipulate the object as seen in figure 5.12. 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 interact with the ring and manipulate it. It is possible that some degree of the proteus effect [234] 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 performing the peg-transfer task, confirming 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 5.13. 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.

5.5.3 Movement Economy

The analyses on the economy of motion revealed that users were significantly more economical with their movement of the manipulated object (ring) when provisioned with the Avatar representation more so than the Avatar-Gain and No-Avatar representations as evinced in figure 5.14. The heightened movement distance of the ring with the Avatar-Gain representation is understandable given the translational offset that is added to users' end effectors in this representation causing a visuo-propriceptive mismatch as mentioned prior. Interestingly, the ring's movement distance was also considerably higher without an avatar than with a co-located avatar, but not different from the gain-based representation. These findings hence refuted hypothesis H8. Given that this economy of motion was measured as a function of the distance moved when the ring was grasped or held, there should not have been any differences between the No-Avatar and Avatar representations. It is unclear as to why this effect was observed and further research is required to ascertain if behavioral aspects or individual differences can help explain these observations. It is also possible that the auditory feedback of the tracked end effector collisions with the pegs when not provisioned with an avatar (No-Avatar) could have resulted in users hearing collisions when their actual hands were not colliding, thus causing more movement. The findings with respect to the Avatar-Gain representations are supported by the findings obtained by other research investigations suggesting that a visuo-proprioceptive mismatch that is concomitant with the Avatar-Gain representation causes users to move the manipulated object more or be less economical in terms of its movement [27, 29].

5.5.4 Subjective Perceptions of Usability and Workload

Results from the subjective questionnaires gathered on users' perceived levels of usability and workload associated with their interactions shed some light on their perceptions and experience in this study. In general, users perceived the interaction system to be more usable and less taxing (workload) when provided with an avatar representation that best approximated the system's tracking of their actual hands or the 'Avatar' representation in this study. In contrast, users found the Avatar-Gain representation to be less usable (higher PSSUQ score) and perceived a higher overall workload when using the same, thus partially supporting hypothesis H9. These results validate the results obtained from the previous study suggesting that visualizing the interacting layer corresponding to users' actual hands is highly beneficial from a user experience perspective. While it might have been beneficial to additionally investigate a fourth avatar representation featuring the same translation gain without the avatar's visualization, the lack of generalizability and applicability of the potential findings associated with this representation dissuaded the pursuit of this investigation. It is, however, easy to see how not provisioning this visualization would degrade performance and score lower on aspects of usability and workload in comparison to the Avatar-Gain representation. For these reasons, we strongly recommend visualizing the interacting layer to significantly improve the user experience in mixed reality interactions.

While we did not observe any main effect of the interaction technique, we discovered a fascinating interaction effect between the avatar representation and the interaction technique used to manipulate the ring (figures 5.16 and 5.15). With an avatar, there doesn't seem to be much of a difference between using the Pinch-to-grasp or Stick-on-touch technique from a usability and workload standpoint. However, without an avatar, users find object interactions with pinching to be significantly more usable and less workload-inducing than a sticky finger approach. These results together offer partial support for hypothesis H10. It hence comes as no surprise that contemporary mixed reality devices continue to facilitate interactions with objects 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 tend to take less time when using the latter technique, they are less accurate and more than anything, perceive this form of interaction as less usable and more taxing.

Taking all the findings together, it seems appropriate for mixed reality 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 in the experience. It is equally important to understand the consequences of these representations and different interaction techniques on different aspects of the experience apart from just interaction performance. 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 associated with near-field interactions, including accuracy, efficiency, and usability. In terms of physicality, a decreased salience of the visual information of relevance associated with purely physical components makes it challenging to perform interactions when the interacting components involve both virtual and physical components. If accuracy is crucial, it behooves designers to visualize the interacting layer corresponding to the registered positions of physical components that are central to the experience. In contrast, when the interacting entities are both virtual, there seems to be a concordance that affords superior interaction performance. With respect to the interaction technique, there appear to be different situations that merit either a pinch-based or stick-based approach. The pinch interaction technique may be more suitable when higher levels of accuracy are desired but not necessarily for efficiency. In contrast, the stick technique seems to lend itself more towards improvements in speed while trading off accuracy. The results obtained in this work hence tend to 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. Overall this work serves to show that virtual end effector representations strongly influence how well users are able to interact in mixed reality experiences.

5.6 Limitations

A number of findings obtained as a result of this study must be qualified based on the registration and tracking performance associated with the system. While the calibration routine employed in the experiment allowed for the careful registration of the pegs on the physical peg board to the best degree possible, it must be noted that the maximum precision attainable was 1mm and 0.1 degrees for positional and orientational references respectively, implying that the tracking system is not flawless. Secondly, the physical apparatus that was used in the study was built using a wooden pegboard and wooden pegs. Unless these physical components were fabricated using metallic substances, their dimensions and scales can only be so accurate. While 3D printed prototypes were initially developed and tested, challenges with regard to printing standalone pegs without a support structure deterred us from pursuing this option. Lastly, the design employed in this study required users to continually switch avatar representations after each phase which comprised 21 trials. This may have limited an in-depth investigation of the calibration of performance associated with the Avatar-Gain representation.

5.7 Conclusion and Future Work

In this study, we empirically examined the impact of provisioning different augmented end-effector representations on users' performance in a near-field object manipulation-based peg-transfer task in a mixed reality setting. Users were tasked with carefully transferring a holographic ring from one destination peg to another for a number of trials while avoiding collisions with the pegs. A multi-factorial design was employed, investigating three factors: the physicality of the pegs (three variations), the augmented avatar representation (three variations), and the interaction technique used to manipulate the ring (two variations). The physicality of the pegs was manipulated as a between-subjects factor across three experimental conditions, namely, Physical (no holograms overlaid on physical components), Virtual+Physical (holograms overlaid on the physical components), and Virtual (only virtual holograms without physical components). The avatar representation and interaction technique, however, were varied as within-subjects factors across multiple phases each consisting of an equal number of trials. A balanced Latin square design was used within each physicality level to eliminate order effects. Results of the study suggest that users were significantly more accurate in the task 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 without gains tend to significantly improve performance. Settling on an interaction technique to manipulate objects would depend on whether accuracy or efficiency is a priority for these tasks. 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 interaction. Furthering this idea, it would be interesting to look at how the properties like the rigidity, roughness/smoothness, pliancy, and hardness of these objects affect such mixed reality interactions. It remains to be seen if partial visualization of the interacting layers associated with registered and tracked objects would be sufficient enough to see an improvement in performance and calibration. The photorealism associated with the rendering of virtual components is yet another factor that could moderate the results. Another area that is ripe for investigation revolves around comparing techniques like ray-casting against gain-based representations on users' ability to calibrate performance over time. Investigating all these phenomena using video see-through mixed reality experiences potentiates a plethora of avenues for research that future technologists and innovators will collectively draw a richness of wealth and knowledge from.

Chapter 6

Contributions

User interaction continues to remain a central component of the user experience with respect to extended reality technologies wherein users interact with virtual integrants. Of the many depth or space regions in which these interactions take place in these mediums, near-field interactions with virtual objects subsume many, if not most, of the interactions that tend to occur as a result of the technology's utilization. Users interact in the near-field by manipulating and retrieving objects. Doing so relies on the ability to effectively coordinate skills, including hand-eye coordination, bilateral integration, and fine motor control, thus requiring precise perception-action coordination from an interaction standpoint. Providing users with a visual representation of themselves in the medium they are currently immersed in allows them to be able to interact effectively with virtual entities. These user representations provide users with a frame of reference depicting a spatial correspondence of the user within the medium and are commonly referred to as self-avatars in virtual, augmented, and mixed reality. Given the challenges associated with provisioning tracked full-body user representations, it is common only to represent users' end-effectors, or simply stated, the distal tools or entities that come into contact with the object(s) being manipulated. Through a series of investigations, this dissertation evaluates the effects of virtual end effector representations on near-field object retrieval interactions in XR settings. Through studies conducted in virtual, augmented, and mixed reality, conclusions about the virtual representation of end-effectors are drawn, and inferences are made for the future of near-field interaction in XR to draw upon from.

6.1 Intellectual Merit

In the initial study, the effects of virtual end-effector representations on near-field interactions were investigated in a fully immersive virtual reality setting. Specifically, their impact on the perceptions of a dynamic affordance-based object retrieval task was examined by conducting an experiment in which participants repeatedly retrieved target objects from within a box while avoiding a moving threat. The results of the study show that when representing users' end-effector as a hand, a discordant mapping between the input modality used to perform the task and the visual end-effector representation does indeed affect interaction performance. This aligns with research showing that different virtual hand representations affect how affordances are perceived in immersive virtual reality [95], and extends findings that suggest that the type of end-effector representation provided affects interaction performance in static settings [130] to also include dynamic contexts. Furthermore, this discordant mapping leads to an increased workload and a diminished ability to calibrate performance in scenarios involving virtual obstacles whose spatio-temporal dynamics continuously change. We also note that when representing the end-effector as a human-like hand, the subjective perceived sense of embodiment is more, even if it comes at the cost of performance.

As an initial foray into the representation of users' end-effectors (hands) in augmented reality, a completely different medium from immersive virtual reality, we conceptualized the second study largely inspired by the relatively recent and renewed promise shown by avatarization in this medium. In order to be able to empirically investigate this phenomenon, a custom end-effector visualization mechanism was developed using a software-driven camera vision-based approach. This constituted the realization of a platform that allows users' actual hands to be tracked, and further visualized in real-time in an optical see-through head-mounted display. As with any AR system, the tracking module or interaction layer is actually responsible for interactions that take place [67]. The novelty of this custom visualization module lies in the ability to visually depict the user's tracked hands as an avatar which precisely and accurately represents the interacting layer of the system. In this study, the effects of provisioning users with a visualized representation of their tracked end-effector were evaluated in a near-field interaction scenario. Users were repeatedly tasked with retrieving an augmented virtual target from within a field of non-target obstacles in a virtual box while avoiding collisions with the obstacles and the box. The findings obtained as a result suggest that avatarizing (visualizing) users' end-effectors significantly improves interaction performance due to the visual representation of the task-dependent interacting layer. In provisioning users with a self-avatar, the visual information based on which the interaction task is defined and dependent upon (tracked end-effector) allows users to distalize to the end-effector representation (virtual tool) visualized as a result. This idea is directly supported by research which suggests that in addition to extending the body schema, distalization of the end-effector from the hand to a tool that is currently in use, is critical to allow one to exploit the functional capacity associated with the use of the tool in itself [6]. It was also noted that when provisioned with a visualized end-effector, users' perceived visibility of their actual physical hands was affected negatively. Whether or not this decreased visibility associated with users' real hands is beneficial from an interaction standpoint hinges on whether or not the task-dependent end-effector is visualized. When visualized, it is important to make the avatar or end effector more salient. On the contrary, without the provisioning of an avatar, it is highly important to make one's real hand more salient.

In a third study, we probed into the problem space of interactions occurring within the mixed reality portion of the XR continuum. We specifically created a scenario that allowed for the investigation of how the physical characteristics of interacting entities affect interactions taking place with a virtual hologram. We conceptualized a peg-transfer task wherein users had to carefully move a holographic ring from one peg to another for a number of trials while trying to be as accurate and as efficient as possible. Using a multi-factorial study design, we varied the physicality of the pegs (3 levels) as a between-subjects factor and manipulated the self-avatar-representation (3 avatar representations) and interaction techniques (2 techniques) as within-subjects factors. One of the avatar representations featured a customized avatar with a translational gain, capable extend a user's personal reach envelope. These manipulations were investigated both from the perspectives of performance-related objective outcomes as well as subjectively perceived levels of usability. Results indicated that a decreased salience of the task-relevant information when interacting with purely physical components makes it challenging to perform interactions when the other interacting components are virtual. In terms of the avatar representations, it was found that a co-located self-avatar produced superior performance than both an offset avatar and a non-visualized avatar, highlighting the benefits of visualizing the end-effector and thus the interacting layer. Our findings also indicate that interaction techniques have different trade-offs associated with them in terms of accuracy and efficiency, encouraging application designers to strongly consider the qualities desired of the experience being developed. Along these lines, sticky-finger interactions allow for faster interactions but are not as accurate as pinch-based interactions. From a subjective user-experience perspective, co-located self-avatars tend to be perceived as the most usable while gain-based representations - although useful - tend to be construed as taxing and less usable. Overall, our findings inform mixed reality developers of the need to visualize the components that are physical, especially if these physical components are a part of the interactions taking place with virtual content. To the best of our knowledge, this is the first study conducted to determine the independent and combined effects of the physicality of interacting components, the self-avatar representations, and the interaction techniques used in performing near-field fine-motor interactions in mixed reality settings.

6.2 Broader Impact

It stands to reason that the heritors of knowledge, as a result of scrutinizing this research, will greatly benefit from understanding and knowing how near-field interactions are to be designed while keeping end-effector representations in mind. In a general sense, object manipulation interactions with virtual artifacts in the near field will continue to be of utmost importance, seeing as how XR is already being used in a multitude of settings, including education, training, therapy, and rehabilitation, apart from just gaming. The designers of VR, AR, and MR systems can make inferences about the choices to be made with respect to how users' end-effectors need to be visualized to maximize the efficiency and performance associated with near-field user interactions.

The field of prosthesis is another highly applicable domain that the findings from this dissertation directly contribute to. Human beings may lose body parts or limbs through trauma, disease, or even a congenital disorder. An artificial device or a prosthetic implant allows for the replacement of the missing body part with an intention to restore normal functions of the missing part. As XR continues to show promise for people requiring such implants, the implications drawn from these studies aid in prosthetic design. Another vital aspect within this realm involves training with respect to these implants. Research suggests that great mental effort is required during the first stages of training, making users that require these implants give up very soon [111]. Using XR supports the learning of how one can and should use these virtual implants efficiently before using their physical counterparts. Thus, choosing the right end-effector to be implanted in order to maximize interaction capabilities for these individuals can be inferred from the results of this dissertation. Moreover, the findings obtained as a result of the works investigated in this dissertation may aid designers in choosing the right medium for training (VR vs AR vs MR) such individuals.

Industrial training applications are often designed in XR to train users and workers about safety and efficiency with respect to operating on machine equipment in fast-paced and potentially dangerous environments. It bodes well to draw upon the findings obtained from this research when designing such simulations to allow users to efficiently interact with the near-field objects that constitute these experiences. Moreover, with several operations increasingly relying on remote interaction using XR, it is imperative to design these systems with the knowledge of how to represent users' end-effectors given that these virtual representations are synced with the industrial robots that establish physical control in these situations.

The scope of the upshot of this research also directly informs the design of XR systems that will facilitate remote surgery. Developers of these systems will be able to draw from the findings of each study to design their simulations to account for more accurate near-field interactions requiring precise perception-action coordination. Surgeons, physicians, and doctors that use these systems will also be able to precisely and effectively act on patients and their bodies using the insights gained as a result of this body of work.

Overall, our studies aim to assist XR system designers in creating more effective and efficient interactions with virtual entities. As XR technology continues to be integrated into areas like training, rehabilitation, education, and therapy, significant improvements are expected in applications used in these fields. Consequently, this will positively impact various domains such as architecture, sports, and education, thus benefiting society as a whole. Our experimental activities focus on exploring the impact of embodied interaction through end effector representations (self-avatars) on near field interactions. Additionally, this works helps in understanding how adaptation via calibration can overcome perceptual limitations in the presence of multi-sensory feedback during perceptually-guided manipulation interactions in interactive XR applications.

Appendices

Appendix A   Informed Consent

Information about Being in a Research Study
Clemson University

Object Retrieval from a Virtual Box in Virtual Reality

KEY INFORMATION ABOUT THE RESEARCH STUDY

Voluntary Consent: Dr. Sabarish Babu is inviting you to volunteer for a research study. Dr. Babu is a tenured professor at Clemson University conducting the study with his graduate students at Clemson University.

You may choose not to take part and you may choose to stop taking part at any time. You will not be punished in any way if you decide not to be in the study or to stop taking part in the study. If you choose to stop taking part in this study, the information you have already provided will be used in a confidential manner.

Study Purpose: The purpose of this study is to explore users' accuracy in being able to retrieve an object from a virtual box.

Activities and Procedures: Your part in the study will be to:
1. Answering questions about your demographics.
2. Retrieving objects from a virtual box that features sliding doors that close and open periodically.
3. Wearing a Virtual Reality/Mixed Reality Headset (HTC Vive Pro)
It will take you about 1 hour to complete this study.

Participation Time: It will take you about one hour to be in this study.

Risks and Discomforts:
There are certain risks or discomforts that you might expect if you take part in this research and it involves wearing Virtual Reality headsets. They include minor eye strain or the experience of "simulator sickness", a form of motion sickness. To minimize this risk, we will follow all known practices to reduce the possibility of experiencing simulator sickness. Other labs that have used VR simulators have suspected that nausea may be more likely in participants who have a history of motion sickness. If you have a history of motion sickness, you may not want to participate in this research study. If you believe that you are susceptible to simulator-related nausea or have a history of epilepsy, we recommend that you not participate in this research.

Possible Benefits: We do not know of any way you would directly benefit from taking part in this study. However, this research may expose to novel technologies such as virtual, augmented and mixed reality.

EXCLUSION/INCLUSION REQUIREMENTS

If you have any known history of epilepsy or vertigo, we advise that you refrain from taking part in this study.

To take part in this study you need to have completed Clemson University's COVID-19 Screener test and reporting procedures, and must have tested negative for COVID-19 with a valid permission from Clemson University to access the campus facilities.

Additionally, this study is focusing on right handed individuals, and if you do not self-identify as a right handed individual or are not right hand dominant, we advise to refrain from participating in this study. Participants should have 20/20 vision or corrected 20/20 vision using contact lenses to participate in the study.

INCENTIVES

Participants recruited from the Department of Psychology Subject Pool (SONA). Upon completion of the study, you will receive 4 credits for participating in this study. No partial credit will be offered, you will have to complete all the study activities to receive the 4 credits.

AUDIO/VIDEO RECORDING AND PHOTOGRAPHS

No audio nor video recordings will be collected in this study.

EQUIPMENT AND DEVICES THAT WILL BE USED IN THE RESEARCH STUDY

Although highly unlikely, if you happen to feel uncomfortable in any way (dizzy, lightheaded, or nauseous) while using the HTC Vive Pro head mounted display, notify the research team immediately. If you continue to experience any discomforts after the study, please contact your preferred healthcare provider and notify the research team.

If you believe that you are susceptible to simulator-related nausea, are pregnant or have a history of epilepsy, you may not want to participate in this research study. Older participants experience simulator related side effects more frequently than university students do. In order to help identify any simulator related side effects, you will answer questions about how you feel after every session.

Notify the research team immediately if you experience any discomforts during the study. If you continue to experience any discomforts after the study, contact your preferred healthcare provider and notify the research team.

PROTECTION OF PRIVACY AND CONFIDENTIALITY

No identifying data will be stored, and any data collected will be kept in locked cabinets or password-protected computers to which only researchers associated with this study have access. Any physical documents associated with this study will be stored in a locked cabinet. The results of this experiment will be summarized across all participants, and no information specifically identifying you will be presented. Your identity will not be revealed in any publication that might result from this study. All data will be stored securely in password protected folders for the duration of study and for a period of 3 years after the study, after which it will be deleted. Data from this study will not be used in future studies, and will not be shared with other researchers who are not a part of this protocol.

We might be required to share the information we collect from you with the Clemson University Office of Research Compliance and the federal Office for Human Research Protections. If this happens, the information will only be used to find out if we ran this study properly and protected your rights in the study.

CONTACT INFORMATION

If you have any questions or concerns about your rights in this research study, please contact the Clemson University Office of Research Compliance (ORC) at 864-656-0636 or irb@clemson.edu. If you are outside of the Upstate South Carolina area, please use the ORC's toll-free number, 866-297-3071. The Clemson IRB will not be able to answer some study-specific questions. However, you may contact the Clemson IRB if the research staff cannot be reached or if you wish to speak with someone other than the research staff.

If you have any questions or concerns about this study or if any problems arise, please contact Roshan Venkatakrishnan at Clemson University at rvenkat@clemson.edu.

CONSENT

By signing this consent form, you indicate that you have read the information written above, are at least 18 years of age, have been allowed to ask any questions, and you are voluntarily choosing to take part in this research.

Participant's signature: ____________________________________ Date: __________

Print name: ____

A copy of this form will be given to you.

Appendix B NASA Task Load Index Questionnaire

Subject ID: ____________________________ **Task ID:** ______________________________

RATING SHEET

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

Low High

FRUSTRATION

Low High

For the following questions, choose the factor that represents the more important contributor to the workload associated with the task.

Effort or **Performance**	**Temporal Demand** or **Frustration**
Temporal Demand or **Effort**	**Physical Demand** or **Frustration**

Performance or **Frustration**	**Physical Demand** or **Temporal Demand**
Physical Demand or **Performance**	**Temporal Demand** or **Mental Demand**

Frustration or **Effort**	**Performance** or **Mental Demand**
Performance or **Temporal Demand**	**Mental Demand** or **Effort**

Mental Demand or **Physical Demand**	**Effort** or **Physical Demand**
Frustration or **Mental Demand**	

Appendix C Embodiment Questionnaire

Embodiment Survey

I felt out of my body

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt as if my (real) hand was drifting toward the virtual hand or as if the virtual hand was drifting toward my (real) hand

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt as if the movements of the virtual hand were influencing my own hand's movements

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

It felt as if my (real) hand was turning into the virtual hand

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

At some point it felt as if my real hand was starting to take on the posture or shape of the virtual hand that I saw

never	almost	rarely	half of the	often	most of the	always

○	never ○	○	time ○	○	time ○	○

I felt like my virtual hand looks different from how a real human hand would look

Strongly Disagree	Disagree	Moderately Disgree	Neutral	Moderately Agree	Agree	Strongly Agree
○	○	○	○	○	○	○

I felt as if my real hand had changed

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt a sensation of hesitation and care in my hand when I was about to collide with the non target balls and walls.

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt that my own hand could be affected by the non target balls and walls

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt as if the virtual hand was my hand

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

At some point it felt that the virtual hand resembled my own (real) hand, in terms of shape, size or other visual features

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt as if my hand was located where I saw the virtual hand

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt like I could control the virtual hand as if it was my own real hand

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I felt like the virtual hand looked and behaved like how my real hand does

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

I seemed as though collisions were caused by my virtual hand rather than my own real hand

Strongly Disagree	Disagree	Moderately Disagree	Neutral	Moderately Agree	Agree	Strongly Agree
○	○	○	○	○	○	○

It seemed as if my hand was retrieving the target balls

never	almost never	rarely	half of the time	often	most of the time	always
○	○	○	○	○	○	○

Appendix D Post Study System Questionnaire

PSSUQ Survey

Please answer the following questions with respect to the interactions with the system:

Overall, I am satisfied with how easy it is to use this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

It was simple to use this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

I was able to complete the tasks and scenarios quickly using this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

I felt comfortable using this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

It was easy to learn to use this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

I believe I can be productive with this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

The system clearly indicates how to fix issues/errors/problems.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

Whenever I made a mistake using the system, I could recover easily and quickly.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

The location of where the system thinks my hand is was clear.

Strongly	Agree	Somewhat	Neutral	Somewhat	Disagree	Strongly

Agree	○	Agree	○	Disagree	○	Disagree
○		○		○		○

It was easy to find out where the system located my hands.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

The system was effective in helping me complete the tasks and scenarios.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

It was clear how the tracking of the system worked.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

Using this system was pleasant.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

I liked using this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

This system has all the functions and capabilities I expect it to have.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

Overall, I am satisfied with this system.

Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○	○

Appendix E Balanced Latin Square Design

PID	Condition	Phase1	Phase2	Phase3	Phase4	Phase5	Phase6
1	Virtual	Pin-NoAv	Pin-Av	Pin-AvG	Stick-NoAv	Stick-Av	Stick-AvG
2	Virtual	Pin-NoAv	Pin-AvG	Pin-Av	Stick-NoAv	Stick-AvG	Stick-Av
3	Virtual	Pin-AvG	Pin-NoAv	Pin-Av	Stick-AvG	Stick-NoAv	Stick-Av
4	Virtual	Stick-AvG	Stick-Av	Stick-NoAv	Pin-AvG	Pin-Av	Pin-NoAv
5	Virtual	Stick-Av	Stick-AvG	Stick-NoAv	Pin-Av	Pin-AvG	Pin-NoAv
6	Virtual	Stick-Av	Stick-NoAv	Stick-AvG	Pin-Av	Pin-NoAv	Pin-AvG
7	Virtual	Stick-NoAv	Stick-Av	Stick-AvG	Pin-NoAv	Pin-Av	Pin-AvG
8	Virtual	Stick-NoAv	Stick-AvG	Stick-Av	Pin-NoAv	Pin-AvG	Pin-Av
9	Virtual	Stick-AvG	Stick-NoAv	Stick-Av	Pin-AvG	Pin-NoAv	Pin-Av
10	Virtual	Pin-AvG	Pin-Av	Pin-NoAv	Stick-AvG	Stick-Av	Stick-NoAv
11	Virtual	Pin-Av	Pin-AvG	Pin-NoAv	Stick-Av	Stick-AvG	Stick-NoAv
12	Virtual	Pin-Av	Pin-NoAv	Pin-AvG	Stick-Av	Stick-NoAv	Stick-AvG
13	Physical	Pin-NoAv	Pin-Av	Pin-AvG	Stick-NoAv	Stick-Av	Stick-AvG
14	Physical	Pin-NoAv	Pin-AvG	Pin-Av	Stick-NoAv	Stick-AvG	Stick-Av
15	Physical	Pin-AvG	Pin-NoAv	Pin-Av	Stick-AvG	Stick-NoAv	Stick-Av
16	Physical	Stick-AvG	Stick-Av	Stick-NoAv	Pin-AvG	Pin-Av	Pin-NoAv
17	Physical	Stick-Av	Stick-AvG	Stick-NoAv	Pin-Av	Pin-AvG	Pin-NoAv
18	Physical	Stick-Av	Stick-NoAv	Stick-AvG	Pin-Av	Pin-NoAv	Pin-AvG
19	Physical	Stick-NoAv	Stick-Av	Stick-AvG	Pin-NoAv	Pin-Av	Pin-AvG
20	Physical	Stick-NoAv	Stick-AvG	Stick-Av	Pin-NoAv	Pin-AvG	Pin-Av
21	Physical	Stick-AvG	Stick-NoAv	Stick-Av	Pin-AvG	Pin-NoAv	Pin-Av
22	Physical	Pin-AvG	Pin-Av	Pin-NoAv	Stick-AvG	Stick-Av	Stick-NoAv
23	Physical	Pin-Av	Pin-AvG	Pin-NoAv	Stick-Av	Stick-AvG	Stick-NoAv
24	Physical	Pin-Av	Pin-NoAv	Pin-AvG	Stick-Av	Stick-NoAv	Stick-AvG
25	V+P	Pin-NoAv	Pin-Av	Pin-AvG	Stick-NoAv	Stick-Av	Stick-AvG
26	V+P	Pin-NoAv	Pin-AvG	Pin-Av	Stick-NoAv	Stick-AvG	Stick-Av
27	V+P	Pin-AvG	Pin-NoAv	Pin-Av	Stick-AvG	Stick-NoAv	Stick-Av
28	V+P	Stick-AvG	Stick-Av	Stick-NoAv	Pin-AvG	Pin-Av	Pin-NoAv
29	V+P	Stick-Av	Stick-AvG	Stick-NoAv	Pin-Av	Pin-AvG	Pin-NoAv
30	V+P	Stick-Av	Stick-NoAv	Stick-AvG	Pin-Av	Pin-NoAv	Pin-AvG
31	V+P	Stick-NoAv	Stick-Av	Stick-AvG	Pin-NoAv	Pin-Av	Pin-AvG
32	V+P	Stick-NoAv	Stick-AvG	Stick-Av	Pin-NoAv	Pin-AvG	Pin-Av
33	V+P	Stick-AvG	Stick-NoAv	Stick-Av	Pin-AvG	Pin-NoAv	Pin-Av
34	V+P	Pin-AvG	Pin-Av	Pin-NoAv	Stick-AvG	Stick-Av	Stick-NoAv
35	V+P	Pin-Av	Pin-AvG	Pin-NoAv	Stick-Av	Stick-AvG	Stick-NoAv
36	V+P	Pin-Av	Pin-NoAv	Pin-AvG	Stick-Av	Stick-NoAv	Stick-AvG

Bibliography

[1] Alex Adkins, Lorraine Lin, Aline Normoyle, Ryan Canales, Yuting Ye, and Sophie Jörg. Evaluating grasping visualizations and control modes in a vr game. *ACM Transactions on Applied Perception (TAP)*, 18(4):1–14, 2021.

[2] Maadh Al-Kalbani, Ian Williams, and Maite Frutos-Pascual. Analysis of medium wrap freehand virtual object grasping in exocentric mixed reality. In *2016 IEEE international symposium on mixed and augmented reality (ISMAR)*, pages 84–93. IEEE, 2016.

[3] Maadh Al-Kalbani, Ian Williams, and Maite Frutos-Pascual. Improving freehand placement for grasping virtual objects via dual view visual feedback in mixed reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 279–282, 2016.

[4] John Aliprantis, Markos Konstantakis, Rozalia Nikopoulou, Phivos Mylonas, and George Caridakis. Natural interaction in augmented reality context. In *VIPERC@ IRCDL*, pages 50–61, 2019.

[5] Ayman Alzayat, Mark Hancock, and Miguel A Nacenta. Quantitative measurement of tool embodiment for virtual reality input alternatives. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.

[6] Michael A Arbib, James B Bonaiuto, Stéphane Jacobs, and Scott H Frey. Tool use and the distalization of the end-effector. *Psychological Research PRPF*, 73(4):441–462, 2009.

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

[8] Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE virtual reality (VR)*, pages 3–10. IEEE, 2016.

[9] Thomas Aslop. Extended reality(XR): AR, VR, and MR - statistics & facts, 2022. `https://www.statista.com/topics/6072/extended-reality-xr/#topicHeader__wrapper`.

[10] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pages 1968–1979, 2016.

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

[12] Sriram Karthik Badam, Arjun Srinivasan, Niklas Elmqvist, and John Stasko. Affordances of input modalities for visual data exploration in immersive environments. In *2nd Workshop on Immersive Analytics*, 2017.

[13] Esther Z Barsom, Maurits Graafland, and Marlies P Schijven. Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical endoscopy*, 30(10):4174–4183, 2016.

[14] Anil Ufuk Batmaz, Aunnoy K Mutasim, Morteza Malekmakan, Elham Sadr, and Wolfgang Stuerzlinger. Touch the wall: Comparison of virtual and augmented reality with conventional 2d screen eye-hand coordination training systems. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 184–193. IEEE, 2020.

[15] Anil Ufuk Batmaz and Wolfgang Stuerzlinger. The effect of rotational jitter on 3d pointing tasks. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2019.

[16] Brett Benda, Shaghayegh Esmaeili, and Eric D Ragan. Determining detection thresholds for fixed positional offsets for virtual hand remapping in virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 269–278. IEEE, 2020.

[17] Ayush Bhargava. *The effect of anthropometric properties of self-avatars on action capabilities in virtual reality*. PhD thesis, Clemson University, 2019.

[18] Ayush Bhargava, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Kathryn Lucaites, Hannah Solini, Andrew C Robb, Christopher C Pagano, and Sabarish V Babu. Can i squeeze through? effects of self-avatars and calibration in a person-plus-virtual-object system on perceived lateral passability in vr. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2348–2357, 2023.

[19] Ayush Bhargava, Roshan Venkatakrishnan, Rohith Venkatakrishnan, Hannah Solini, Kathryn Lucaites, Andrew C Robb, Christopher C Pagano, and Sabarish V Babu. Did i hit the door? effects of self-avatars and calibration in a person-plus-virtual-object system on perceived frontal passability in vr. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4198–4210, 2021.

[20] Ayush Bhargava, Roshan Venkatakrishnan, Rohith Venkatakrishnan, Hannah Solini, Kathryn Lucaites, Andrew C Robb, Christopher C Pagano, and Sabarish V Babu. Empirically evaluating the effects of eye height and self-avatars on dynamic passability affordances in virtual reality. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 308–317. IEEE, 2023.

[21] Geoffrey P Bingham and Christopher C Pagano. The necessity of a perception–action approach to definite distance perception: Monocular distance perception to guide reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1):145, 1998.

[22] Olaf Blanke and Thomas Metzinger. Full-body illusions and minimal phenomenal selfhood. *Trends in cognitive sciences*, 13(1):7–13, 2009.

[23] Julia JC Blau and Jeffrey B Wagman. *Introduction to Ecological Psychology: A lawful approach to perceiving, acting, and cognizing*. Taylor & Francis, 2022.

[24] Christoph W Borst and Arun P Indugula. Realistic virtual grasping. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pages 91–98. IEEE, 2005.

[25] Sanne MBI Botden, Sonja N Buzink, Marlies P Schijven, and Jack J Jakimowicz. Augmented versus virtual reality laparoscopic simulation: What is the difference? *World journal of surgery*, 31(4):764–772, 2007.

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

[27] David Brickler and Sabarish 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.

[28] David Brickler, Robert J Teather, Andrew T Duchowski, and Sabarish 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.

[29] David Brickler, Matias Volonte, Jeffrey W Bertrand, Andrew T Duchowski, and Sabarish 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)*, pages 28–37. IEEE, 2019.

[30] Jelle Bruineberg, Ludovic Seifert, Erik Rietveld, and Julian Kiverstein. Metastable attunement and real-life skilled behavior. *Synthese*, 199(5):12819–12842, 2021.

[31] Volkert Buchmann, Trond Nilsen, and Mark Billinghurst. Interaction with partially transparent hands and objects. 2005.

[32] Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy 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*, pages 212–221, 2004.

[33] Martinus Buekers, Gilles Montagne, Aymar de Rugy, and Michel Laurent. The regulation of externally paced human locomotion in virtual reality. *Neuroscience Letters*, 275(3):171–174, 1999.

[34] Ian M Bullock, Thomas Feix, and Aaron 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)*, pages 5768–5772. IEEE, 2015.

[35] Giuseppe Caggianese, Luigi Gallo, and Pietro Neroni. The vive controllers vs. leap motion for interactions in virtual environments: a comparative evaluation. In *International Conference on Intelligent Interactive Multimedia Systems and Services*, pages 24–33. Springer, 2018.

[36] Ryan Canales and Sophie 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*, pages 1–6, 2020.

[37] Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. Virtual grasping feedback and virtual hand ownership. In *ACM Symposium on Applied Perception 2019*, pages 1–9, 2019.

[38] Juan Cristobal Castro-Alonso, Paul Ayres, and Fred Paas. Dynamic visualisations and motor skills. In *Handbook of human centric visualization*, pages 551–580. Springer, 2014.

[39] Yi-Ting Chen, Chi-Hsuan Hsu, Chih-Han Chung, Yu-Shuen Wang, and Sabarish V Babu. Ivrnote: design, creation and evaluation of an interactive note-taking interface for study and reflection in vr learning environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 172–180. IEEE, 2019.

[40] Zhaorui Chen, Jinzhu Li, Yifan Hua, Rui Shen, and Anup Basu. Multimodal interaction in augmented reality. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 206–209. IEEE, 2017.

[41] Kai-Yin Cheng, Rong-Hao Liang, Bing-Yu Chen, Rung-Huei Laing, and Sy-Yen Kuo. icon: utilizing everyday objects as additional, auxiliary and instant tabletop controllers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1155–1164, 2010.

[42] Benjamin J Chihak, Jodie M Plumert, Christine J Ziemer, Sabarish Babu, Timofey Grechkin, James F Cremer, and Joseph K Kearney. Synchronizing self and object movement: how child and adult cyclists intercept moving gaps in a virtual environment. *Journal of experimental psychology: human perception and performance*, 36(6):1535, 2010.

[43] Marina A Cidota, Rory MS Clifford, Stephan G Lukosch, and Mark Billinghurst. Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 172–177. IEEE, 2016.

[44] Marina A Cidota, Stephan G Lukosch, Paulina JM Bank, and PW Ouwehand. Towards engaging upper extremity motor dysfunction assessment using augmented reality games. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 275–278. IEEE, 2017.

[45] Michael E Cinelli, Aftab E Patla, and Fran Allard. Behaviour and gaze analyses during a goal-directed locomotor task. *Quarterly Journal of Experimental Psychology*, 62(3):483–499, 2009.

[46] Rachel O Coats, Jing S Pan, and Geoffrey P Bingham. Perturbation of perceptual units reveals dominance hierarchy in cross calibration. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):328, 2014.

[47] Stefany Coxe, Stephen G West, and Leona 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.

[48] James E Cutting and Peter M Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of space and motion*, pages 69–117. Elsevier, 1995.

[49] Dragoş Datcu, Stephan Lukosch, and Frances Brazier. On the usability and effectiveness of different interaction types in augmented reality. *International Journal of Human-Computer Interaction*, 31(3):193–209, 2015.

[50] Brian Day, Elham Ebrahimi, Leah S Hartman, Christopher C Pagano, Andrew C Robb, and Sabarish V Babu. Examining the effects of altered avatars on perception-action in virtual reality. *Journal of Experimental Psychology: Applied*, 25(1):1, 2019.

[51] Alexandre Gomes de Siqueira, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Ayush Bhargava, Kathryn Lucaites, Hannah Solini, Moloud Nasiri, Andrew Robb, Christopher Pagano, Brygg Ullmer, et al. Empirically evaluating the effects of perceptual information channels on the size perception of tangibles in near-field virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 1–10. IEEE, 2021.

[52] Frédérique De Vignemont. Embodiment, ownership and disownership. *Consciousness and cognition*, 20(1):82–93, 2011.

[53] Dr.meter. Digital light instensity meter: Measure intensity of light using this digital illuminance/light meter, lx1010b, by dr.meter, 2022. `https://drmeter.com/products/digital-lux-meter-digital-illuminance-light-meter-lx1010b-dr-meter`.

[54] Elham Ebrahimi, Leah S Hartman, Andrew Robb, Christopher C Pagano, and Sabarish 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)*, pages 1–8. IEEE, 2018.

[55] Elham Ebrahimi, Andrew Robb, Leah S Hartman, Christopher C Pagano, and Sabarish 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*, pages 1–8, 2018.

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

[57] Shaghayegh Esmaeili, Brett Benda, and Eric 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)*, pages 453–462. IEEE, 2020.

[58] Brett R Fajen. *Visual control of locomotion*. Cambridge University Press, 2021.

[59] Brett R Fajen, Gabriel Diaz, and Christopher Cramer. Reconsidering the role of movement in perceiving action-scaled affordances. *Human Movement Science*, 30(3):504–533, 2011.

[60] Brett R Fajen and Jonathan S Matthis. Direct perception of action-scaled affordances: the shrinking gap problem. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5):1442, 2011.

[61] Alex Olwal Steven Feiner. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proc. UIST*, volume 3, pages 81–82, 2003.

[62] Tiare Feuchtner and Jörg Müller. Extending the body for interaction with reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5145–5157, 2017.

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

[64] Maite Frutos-Pascual, Chris Creed, and Ian Williams. Head mounted display interaction evaluation: manipulating virtual objects in augmented reality. In *IFIP Conference on Human-Computer Interaction*, pages 287–308. Springer, 2019.

[65] Michael J Fu, Andrew D Hershberger, Kumiko Sano, and M Cenk Çavuşoğlu. Effect of visuo-haptic co-location on 3d fitts' task performance. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3460–3467. IEEE, 2011.

[66] Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21, 2000.

[67] Adelaide Charlotte Sarah Genay, Anatole Lécuyer, and Martin Hachet. Being an avatar" for real": a survey on virtual embodiment in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[68] Alexandre Gomes de Siqueira, Ayush Bhargava, Rohith Venkatakrishnan, and Roshan Venkatakrishnan. Ppcards: Toward enhancing electronic prototyping with editions of a card-based platform. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–11, 2021.

[69] Eric J Gonzalez and Sean Follmer. Investigating the detection of bimanual haptic retargeting in virtual reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, pages 1–5, 2019.

[70] Roe Goodman and Nolan R Wallach. *Representations and invariants of the classical groups*. Cambridge University Press, 2000.

[71] Timofey Y Grechkin, Jodie M Plumert, and Joseph K Kearney. Dynamic affordances in embodied interactive systems: The role of display and mode of locomotion. *IEEE transactions on visualization and computer graphics*, 20(4):596–605, 2014.

[72] Taejin Ha, Steven Feiner, and Woontack 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)*, pages 219–228. IEEE, 2014.

[73] Kelly S Hale and Kay M Stanney. *Handbook of virtual environments: Design, implementation, and applications*. CRC Press, 2014.

[74] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[75] Henry Head and Gordon Holmes. Sensory disturbances from cerebral lesions. *Brain*, 34(2-3):102–254, 1911.

[76] Henry Head, William Halse Rivers Rivers, James Col Sherren, Gordon Col Holmes, Theodore Col Thompson, and George Col Riddoch. Studies in neurology, in two volumes, vol 1. 1920.

[77] Chris Heinrich, Matthew Cook, Tobias Langlotz, and Holger Regenbrecht. My hands? importance of personalised virtual hands in a neurorehabilitation scenario. *Virtual Reality*, 25(2):313–330, 2021.

[78] William LeRoy Heinrichs, Patricia Youngblood, Phillip M Harter, and Parvati Dev. Simulation for team training and assessment: case studies of online training with virtual worlds. *World journal of surgery*, 32(2):161–170, 2008.

[79] Naoya Hirose. An ecological approach to embodiment and cognition. *Cognitive Systems Research*, 3(3):289–299, 2002.

[80] Thuong N Hoang, Ross T Smith, and Bruce H Thomas. Ultrasonic glove input device for distance-based interactions. In *2013 23rd International Conference on Artificial Reality and Telexistence (ICAT)*, pages 46–53. IEEE, 2013.

[81] Simon Hoermann, Elizabeth A Franz, and Holger Regenbrecht. Referred sensations elicited by video-mediated mirroring of hands. *PLoS One*, 7(12):e50942, 2012.

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

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

[84] Sarah Howlett, Richard Lee, and Carol O'Sullivan. A framework for comparing task performance in real and virtual scenes. In *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*, pages 119–122, 2005.

[85] Chi-Hsuan Hsu, Chi-Han Chung, Rohith Venkatakrishnan, Roshan Venkatakrlshnan, Yu-Shuen Wang, and Sabarish V Babu. Comparative evaluation of digital writing and art in real and immersive virtual environments. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 1–10. IEEE, 2021.

[86] Yi-Hsuan Huang, Roshan Venkatakrishnan, Rohith Venkatakrishnan, Sabarish V Babu, and Wen-Chieh Lin. Using audio reverberation to compensate distance compression in virtual reality. In *ACM Symposium on Applied Perception 2021*, pages 1–10, 2021.

[87] Igroup. Igroup Presence Questionnaire (IPQ), 2018. `http://www.igroup.org/pq/ipq/`.

[88] Sylvia Irawati, Scott Green, Mark Billinghurst, Andreas Duenser, and Heedong Ko. " move the couch where?": developing an augmented reality multimodal interface. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 183–186. IEEE, 2006.

[89] Shaziela Ishak, Karen E Adolph, and Grace C Lin. Perceiving affordances for fitting through apertures. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6):1501, 2008.

[90] Hidetoshi Ishiyama and Shuichi 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)*, pages 187–188. IEEE, 2016.

[91] Shubham Jain, Thomas Schweiss, Simon Bender, and Dirk Werth. Omnichannel retail customer experience with mixed-reality shopping assistant systems. In *International Symposium on Visual Computing*, pages 504–517. Springer, 2021.

[92] Jacek Jankowski and Martin Hachet. A survey of interaction techniques for interactive 3d environments. In *Eurographics 2013-STAR*, 2013.

[93] Roland S Johansson, Göran Westling, Anders Bäckström, and J Randall Flanagan. Eye–hand coordination in object manipulation. *Journal of neuroscience*, 21(17):6917–6932, 2001.

[94] Adrian S Johnson and Yu Sun. Spatial augmented reality on person: Exploring the most personal medium. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 169–174. Springer, 2013.

[95] Tugce Joy, Emre Ugur, and Inci Ayhan. Trick the body trick the mind: avatar representation affects the perception of available action possibilities in virtual reality. *Virtual Reality*, 26(2):615–629, 2022.

[96] Eunice Jun, Jeanine K Stefanucci, Sarah H Creem-Regehr, Michael N Geuss, and William B Thompson. Big foot: Using the size of a virtual foot to scale gap width. *ACM Transactions on Applied Perception (TAP)*, 12(4):1–12, 2015.

[97] Ed Kaiser, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Xiaoguang Li, Phil Cohen, and Steven Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 12–19, 2003.

[98] Fuminari Kaneko, Keiichiro Shindo, Masaki Yoneta, Megumi Okawada, Kazuto Akaboshi, and Meigen 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.

[99] Maryam Khademi, Hossein Mousavi Hondori, Lucy Dodakian, Steve Cramer, and Cristina V Lopes. Comparing "pick and place" task in spatial augmented reality versus non-immersive virtual reality for rehabilitation setting. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4613–4616. IEEE, 2013.

[100] Ahmed S Khalaf, Sultan A Alharthi, Igor Dolgov, and Z O Toups. A comparative study of hand gesture recognition devices in the context of game design. In *Proceedings of the 2019 ACM international conference on interactive surfaces and spaces*, pages 397–402, 2019.

[101] Chaowanan Khundam, Varunyu Vorachart, Patibut Preeyawongsakul, Witthaya Hosap, and Frédéric Noël. A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. In *Informatics*, volume 8, page 60. Multidisciplinary Digital Publishing Institute, 2021.

[102] Konstantina Kilteni, Raphaela Groten, and Mel Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[103] Luv Kohli, Mary C Whitton, and Frederick P Brooks. Redirected touching: The effect of warping space on task performance. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 105–112. IEEE, 2012.

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

[105] Elena Kokkinara and Mel Slater. Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception*, 43(1):43–58, 2014.

[106] Regis Kopper, Doug A Bowman, Mara G Silva, and Ryan P McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010.

[107] Felix Kosmalla, André Zenner, Marco Speicher, Florian Daiber, Nico Herbig, and Antonio Krüger. Exploring rock climbing in mixed reality environments. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1787–1793, 2017.

[108] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. Augmented reality versus virtual reality for 3d object manipulation. *IEEE transactions on visualization and computer graphics*, 24(2):1038–1048, 2017.

[109] Paul G Kry, Adeline Pihuit, Adrien Bernhardt, and Marie-Paule Cani. Handnavigator: Hands-on interaction for desktop virtual reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 53–60, 2008.

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

[111] Edgard Lamounier, Kenedy Lopes, Alexandre Cardoso, Adriano Andrade, and Alcimar Soares. On the use of virtual and augmented reality for upper limb prostheses training and simulation. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 2451–2454. IEEE, 2010.

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

[113] Kangdon Lee. Augmented reality in education and training. *TechTrends*, 56(2):13–21, 2012.

[114] Minkyung Lee, Mark Billinghurst, Woonhyuk Baek, Richard Green, and Woontack Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Reality*, 17(4):293–305, 2013.

[115] Frank Lentricchia and Thomas McLaughlin. *Critical terms for literary study*. University of Chicago Press, 2010.

[116] James 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.

[117] Markus Leyrer, Sally A Linkenauger, Heinrich H Bülthoff, Uwe Kloos, and Betty Mohler. The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH symposium on applied perception in graphics and visualization*, pages 67–74, 2011.

[118] Jialei Li, Isaac Cho, and Zachary Wartell. Evaluation of cursor offset on 3d selection in vr. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*, pages 120–129, 2018.

[119] Ke Li, Raviraj Nataraj, Tamara L Marquardt, and Zong-Ming Li. Directional coordination of thumb and finger forces during precision pinch. *PloS one*, 8(11):e79400, 2013.

[120] Zong-Ming Li and Jie Tang. Coordination of thumb joints during opposition. *Journal of biomechanics*, 40(3):502–510, 2007.

[121] Lorraine Lin, Aline Normoyle, Alexandra Adkins, Yu Sun, Andrew Robb, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. The effect of hand size and interaction modality on the virtual hand illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 510–518. IEEE, 2019.

[122] Wan-Yi Lin, Ying-Chu Wang, Dai-Rong Wu, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Elham Ebrahimi, Christopher Pagano, Sabarish V Babu, and Wen-Chieh Lin. Empirical evaluation of calibration and long-term carryover effects of reverberation on egocentric auditory depth perception in vr. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 232–240. IEEE, 2022.

[123] Yun-Xuan Lin, Sabarish V Babu, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Ying-Chu Wang, and Wen-Chieh Lin. Towards an immersive virtual simulation for studying cybersickness during spatial knowledge acquisition. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 624–625. IEEE, 2020.

[124] Yun-Xuan Lin, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Elham Ebrahimi, Wen-Chieh Lin, and Sabarish V Babu. How the presence and size of static peripheral blur affects cybersickness in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 17(4):1–18, 2020.

[125] Sally A Linkenauger, Markus Leyrer, Heinrich H Bülthoff, and Betty J Mohler. Welcome to wonderland: The influence of the size and shape of a virtual hand on the perceived size and shape of virtual objects. *PloS one*, 8(7):e68594, 2013.

[126] Hangxin Liu, Zhenliang Zhang, Xu Xie, Yixin Zhu, Yue Liu, Yongtian Wang, and Song-Chun Zhu. High-fidelity grasping in virtual reality using a glove-based system. In *2019 international conference on robotics and automation (icra)*, pages 5180–5186. IEEE, 2019.

[127] Huimin Liu, Zhiquan Wang, Christos Mousas, and Dominic Kao. Virtual reality racket sports: Virtual drills for exercise and training. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 566–576. IEEE, 2020.

[128] Benjamin Lok, Samir Naik, Mary Whitton, and Frederick P Brooks. Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence*, 12(6):615–628, 2003.

[129] Matthew R Longo, Friederike Schüür, Marjolein PM Kammers, Manos Tsakiris, and Patrick Haggard. What is embodiment? a psychometric approach. *Cognition*, 107(3):978–998, 2008.

[130] Christos Lougiakis, Akrivi Katifori, Maria Roussou, and Ioannis-Panagiotis 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)*, pages 510–518. IEEE, 2020.

[131] Gan Lu, Lik-Kwan Shark, Geoff Hall, and Ulrike Zeshan. Immersive manipulation of virtual objects through glove-based hand gesture interaction. *Virtual Reality*, 16(3):243–252, 2012.

[132] Kathryn M Lucaites, Roshan Venkatakrishnan, Ayush Bhargava, Rohith Venkatakrishnan, and Christopher C Pagano. Predicting aperture crossing behavior from within-trial metrics of motor control reliability. *Human Movement Science*, 74:102713, 2020.

[133] Kathryn M Lucaites, Roshan Venkatakrishnan, Rohith Venkatakrishnan, Ayush Bhargava, and Christopher C Pagano. Predictability and variability of a dynamic environment impact affordance judgments. *Ecological Psychology*, 32(2-3):95–114, 2020.

[134] Katie Lucaites, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Ayush Bhargava, and Christopher C Pagano. Perceptions of passability through dynamically moving gaps. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 1941–1942. SAGE Publications Sage CA: Los Angeles, CA, 2019.

[135] Jean-Luc Lugrin, Maximilian Ertl, Philipp Krop, Richard Klüpfel, Sebastian Stierstorfer, Bianka Weisz, Maximilian Rück, Johann Schmitt, Nina Schmidt, and Marc Erich Latoschik. Any "body" there? avatar visibility effects in a virtual reality game. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 17–24. IEEE, 2018.

[136] Anna Macaranas, Alissa N Antle, and Bernhard E Riecke. What is intuitive interaction? balancing users' performance and satisfaction with natural user interfaces. *Interacting with Computers*, 27(3):357–370, 2015.

[137] Madhur Mangalam, Dorothy M Fragaszy, Jeffrey B Wagman, Brian M Day, Damian G Kelty-Stephen, Raoul M Bongers, Dietrich W Stout, and François Osiurak. On the psychological origins of tool use. *Neuroscience & Biobehavioral Reviews*, page 104521, 2022.

[138] Manus. Wearable Manus Gloves, 2022. `https://www.manus-meta.com/products/gloves`.

[139] Mary 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.

[140] Brandon J Matthews, Bruce H Thomas, Stewart Von Itzstein, and Ross T Smith. Remapped physical-virtual interfaces with bimanual haptic retargeting. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 19–27. IEEE, 2019.

[141] Ryan P McMahan, Alexander Joel D Alon, Shaimaa Lazem, Robert J Beaton, David Machaj, Michael Schaefer, Mara G Silva, Anamary Leal, Robert Hagan, and Doug A Bowman. Evaluating natural interaction techniques in video games. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 11–14. IEEE, 2010.

[142] Erin A McManus, Bobby Bodenheimer, Stephan Streuber, Stephan De La Rosa, Heinrich H Bülthoff, and Betty 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*, pages 37–44, 2011.

[143] Michael Meehan, Sharif Razzaque, Mary C Whitton, and Frederick P Brooks. Effect of latency on presence in stressful virtual environments. In *IEEE Virtual Reality, 2003. Proceedings.*, pages 141–148. IEEE, 2003.

[144] David Merrill and Pattie Maes. Augmenting looking, pointing and reaching gestures to enhance the searching and browsing of physical objects. In *International Conference on Pervasive Computing*, pages 1–18. Springer, 2007.

[145] Domenick M Mifsud, Adam S Williams, Francisco Ortega, and Robert 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)*, pages 590–591. IEEE, 2022.

[146] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–292. Spie, 1995.

[147] Mathias Moehring and Bernd Froehlich. Effective manipulation of virtual objects within arm's reach. In *2011 IEEE Virtual Reality Conference*, pages 131–138. IEEE, 2011.

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

[149] Mark Mon-Williams and Geoffrey P Bingham. Calibrating reach distance to visual targets. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3):645, 2007.

[150] Till Nagel and Frank Heidmann. Exploring faceted geo-spatial data with tangible interaction. *GeoViz 2011*, pages 10–11, 2011.

[151] Diederick C Niehorster, Li Li, and Markus 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.

[152] Seungtak Noh, Hui-Shyong Yeo, and Woontack Woo. An hmd-based mixed reality system for avatar-mediated remote collaboration with bare-hand interaction. In *ICAT-EGVE*, pages 61–68, 2015.

[153] Noitom. Noitom Hi5 Glove, 2022. `https://www.noitom.com/hi5-vr-glove`.

[154] Winfried Nöth. Representation in semiotics and in computer science. 1997.

[155] Nami Ogawa, Takuji Narumi, and Michitaka Hirose. Virtual hand realism affects object size perception in body-based scaling. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 519–528. IEEE, 2019.

[156] Nami Ogawa, Takuji Narumi, and Michitaka Hirose. Effect of avatar appearance on detection thresholds for remapped hand movements. *IEEE transactions on visualization and computer graphics*, 27(7):3182–3197, 2020.

[157] Nami Ogawa, Takuji Narumi, Hideaki Kuzuoka, and Michitaka Hirose. Do you feel like passing through walls?: Effect of self-avatar appearance on facilitating realistic behavior in virtual environments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

[158] Elizabeth E O'Neal, Yuanyuan Jiang, Lucas J Franzen, Pooya Rahimian, Junghum Paul Yon, Joseph K Kearney, and Jodie M Plumert. Changes in perception–action tuning over long time scales: How children and adults perceive and act on dynamic affordances when crossing roads. *Journal of experimental psychology: human perception and performance*, 44(1):18, 2018.

[159] Ori Ossmy and Roy Mukamel. Short term motor-skill acquisition improves with size of self-controlled virtual hands. *PloS one*, 12(1):e0168520, 2017.

[160] Riku Otono, Naoya Isoyama, Hideaki Uchiyama, and Kiyoshi 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)*, pages 730–731. IEEE, 2022.

[161] Timothy F O'Connor, Matthew E Fach, Rachel Miller, Samuel E Root, Patrick P Mercier, and Darren J Lipomi. The language of glove: Wireless gesture decoder with low-power and stretchable hybrid electronics. *PloS one*, 12(7):e0179766, 2017.

[162] Christopher C Pagano and Michael T Turvey. Eigenvectors of the inertia tensor and perceiving the orientations of limbs and objects. *Journal of Applied Biomechanics*, 14(4):331–359, 1998.

[163] Tabitha C Peck and Mar Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 1:575943, 2021.

[164] Tabitha C Peck, Sofia Seinfeld, Salvatore M Aglioti, and Mel Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, 22(3):779–787, 2013.

[165] Duc-Minh Pham and Wolfgang Stuerzlinger. Is the pen mightier than the controller? a comparison of input devices for selection in virtual and augmented reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2019.

[166] Catlin Pidel and Philipp Ackermann. Collaboration in virtual and augmented reality: a systematic overview. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 141–156. Springer, 2020.

[167] Wayne Piekarski and Bruce H Thomas. Interactive augmented reality techniques for construction at a distance of 3d geometry. In *Proceedings of the workshop on Virtual environments 2003*, pages 19–28, 2003.

[168] Jeffrey S Pierce, Andrew S Forsberg, Matthew J Conway, Seung Hong, Robert C Zeleznik, and Mark R Mine. Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 39–ff, 1997.

[169] Ivelina V Piryankova, Hong Yu Wong, Sally A Linkenauger, Catherine Stinson, Matthew R Longo, Heinrich H Bülthoff, and Betty J Mohler. Owning an overweight or underweight body: distinguishing the physical, experienced and virtual body. *PloS one*, 9(8):e103428, 2014.

[170] Hanna Fenichel Pitkin. The concept of representation. In *Democracy: A Reader*, pages 155–158. Columbia University Press, 2016.

[171] Thammathip Piumsomboon, David Altimira, Hyungon Kim, Adrian Clark, Gun Lee, and Mark 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)*, pages 73–82. IEEE, 2014.

[172] Thammathip Piumsomboon, Adrian Clark, and Mark 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)*, pages 161–166, 2011.

[173] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*, pages 282–299. Springer, 2013.

[174] Jodie M Plumert and Joseph K Kearney. How do children perceive and act on dynamic affordances in crossing traffic-filled roads? *Child development perspectives*, 8(4):207–212, 2014.

[175] Ronald Poelman, Oytun Akman, Stephan Lukosch, and Pieter Jonker. As if being there: mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 1267–1276, 2012.

[176] Ivan Poupyrev, Tadao Ichikawa, Suzanne Weghorst, and Mark Billinghurst. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum*, volume 17, pages 41–52. Wiley Online Library, 1998.

[177] Michael Prilla, Marc Janßen, and Timo 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.

[178] Dennis R Proffitt and Sally A Linkenauger. Perception viewed as a phenotypic expression. *Action science: Foundations of an emerging discipline*, 171, 2013.

[179] Moritz Quandt, David Hippert, Thies Beinke, and Michael 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.

[180] Steven R Ness, Paul Reimer, Norman Krell, Gabrielle Odowichuck, W Andrew Schloss, and George Tzanetakis. Sonophenology: a tangible interface for sonification of geo-spatial phenological data at multiple time-scales. Georgia Institute of Technology, 2010.

[181] Rafael Radkowski and Christian Stritzke. Interactive hand gesture-based assembly for augmented reality applications. In *Proceedings of the 2012 International Conference on Advances in Computer-Human Interactions*, pages 303–308. Citeseer, 2012.

[182] Holger Regenbrecht, Simon Hoermann, Claudia Ott, Lavell Mueller, and Elizabeth Franz. Manipulating the experience of reality for rehabilitation applications. *Proceedings of the IEEE*, 102(2):170–184, 2014.

[183] Aylen Ricca, Amine Chellali, and Samir Otmane. Influence of hand visualization on tool-based motor skills training in an immersive vr simulator. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 260–268. IEEE, 2020.

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

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

[186] Nina Rosa, Remco C Veltkamp, Wolfgang Hürst, Tanja Nijboer, Carolien Gilbers, and Peter Werkhoven. The supernumerary hand illusion in augmented reality. *ACM Transactions on Applied Perception (TAP)*, 16(2):1–20, 2019.

[187] Sverker Runeson and Gunilla Frykholm. Kinematic specification of dynamics as an informational basis for person-and-action perception: expectation, gender recognition, and deceptive intention. *Journal of experimental psychology: general*, 112(4), 1983.

[188] Christian Sandor, Shinji Uchiyama, and Hiroyuki Yamamoto. Visuo-haptic systems: Half-mirrors considered harmful. In *Second joint eurohaptics conference and symposium on haptic interfaces for virtual environment and teleoperator systems (WHC'07)*, pages 292–297. IEEE, 2007.

[189] Mie Sato, Sota Suzuki, Daiki Ebihara, Sho Kato, and Sato Ishigaki. Pseudo-softness evaluation in grasping a virtual object with a bare hand. In *ACM SIGGRAPH 2016 Posters*, pages 1–1. 2016.

[190] Alexander Schäfer, Gerd Reis, and Didier Stricker. The gesture authoring space: Authoring customised hand gestures for grasping virtual objects in immersive virtual environments. *arXiv e-prints*, pages arXiv–2207, 2022.

[191] Bert Schiettecatte and Jean 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*, pages 3–10, 2008.

[192] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels 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*, pages 1577–1582, 2017.

[193] Sofia Seinfeld, Tiare Feuchtner, Antonella Maselli, and Jörg Müller. User representations in human-computer interaction. *Human–Computer Interaction*, 36(5-6):400–438, 2021.

[194] Sofia Seinfeld, Tiare Feuchtner, Johannes Pinzek, and Jörg Müller. Impact of information placement and user representations in vr on performance and embodiment. *IEEE transactions on visualization and computer graphics*, pages 1–13, 2020.

[195] Dong Woo Seo and Jae Yeol Lee. Direct hand touchable interactions in augmented reality environments for natural and intuitive user experiences. *Expert Systems with Applications*, 40(9):3784–3793, 2013.

[196] Ramiro Serrano, Pedro Morillo, Sergio Casas, and Carolina Cruz-Neira. An empirical evaluation of two natural hand interaction systems in augmented reality. *Multimedia Tools and Applications*, pages 1–27, 2022.

[197] Gordon Simpson, Lucy Johnston, and Michael Richardson. An investigation of road crossing in a virtual environment. *Accident Analysis & Prevention*, 35(5):787–796, 2003.

[198] Paul Skalski, Ron Tamborini, Ashleigh Shelton, Michael Buncher, and Pete Lindmark. Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. *New Media & Society*, 13(2):224–242, 2011.

[199] Tom AB Snijders and Roel J Bosker. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. sage, 2011.

[200] Alcimar Barbosa Soares, Edgard Afonso Lamounier Júnior, Adriano de Oliveira Andrade, and Alexandre 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*, pages 409–426, 2012.

[201] Inês Soares, Ricardo B. Sousa, Marcelo Petry, and António Paulo Moreira. Accuracy and repeatability tests on hololens 2 and htc vive. *Multimodal Technologies and Interaction*, 5(8):47, 2021.

[202] Jean Sreng, Anatole Lécuyer, Christine Mégard, and Claude Andriot. Using visual cues of contact to improve interactive manipulation of virtual objects in industrial assembly/maintenance simulations. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1013–1020, 2006.

[203] Gideon Sroka, Liane S Feldman, Melina C Vassiliou, Pepa A Kaneva, Raad Fayez, and Gerald 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.

[204] Janet L Starkes and Susan Lindley. Can we hasten expertise by video simulations? *Quest*, 46(2):211–222, 1994.

[205] Anthony Steed, Ye Pan, Fiona Zisch, and William Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE virtual reality (VR)*, pages 67–76. IEEE, 2016.

[206] Jeanine K Stefanucci and Michael N Geuss. Big people, little world: The body influences size perception. *Perception*, 38(12):1782–1795, 2009.

[207] Mohamed Suhail, Shyam Prathish Sargunam, Dustin T Han, and Eric D Ragan. Redirected reach in virtual reality: Enabling natural hand interaction at multiple virtual locations with passive haptics. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 245–246. IEEE, 2017.

[208] Sota Suzuki, Haruto Suzuki, and Mie Sato. Grasping a virtual object with a bare hand. In *ACM SIGGRAPH 2014 Posters*, pages 1–1. 2014.

[209] Tanh Quang Tran, HyunJu Shin, Wolfgang Stuerzlinger, and JungHyun Han. Effects of virtual arm representations on interaction in virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pages 1–9, 2017.

[210] Manos Tsakiris, Gita Prabhu, and Patrick Haggard. Having a body versus moving your body: How agency structures body-ownership. *Consciousness and cognition*, 15(2):423–432, 2006.

[211] Michael T Turvey. The primacy of perceiving: An ecological reformulation of perception for understanding memory. *Studies of memory: In honor of Uppsala University's 500th anniversary*, pages 167–222, 1979.

[212] Pier Paolo 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.

[213] Michiel Van Veldhuizen and Xubo Yang. The effect of semi-transparent and interpenetrable hands on object manipulation in virtual reality. *IEEE Access*, 9:17572–17583, 2021.

[214] R Venkatakrishnan, A Robb, and SV Babu. Row vr-can competing with someone make you perform better? 2019.

[215] Rohith Venkatakrishnan, Ayush Bhargava, Roshan Venkatakrishnan, Kathryn M Lucaites, Matias Volonte, Hannah Solini, Andrew C Robb, Christopher Pagano, and Sabarish V Babu. Towards an immersive driving simulator to study factors related to cybersickness. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1201–1202. IEEE, 2019.

[216] Rohith Venkatakrishnan, Roshan Venkatakrishnan, Reza Ghaiumy Anaraky, Matias Volonte, Bart Knijnenburg, and Sabarish V Babu. A structural equation modeling approach to understand the relationship between control, cybersickness and presence in virtual reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 682–691. IEEE, 2020.

[217] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Ayush Bhargava, Kathryn Lucaites, Hannah Solini, Matias Volonte, Andrew Robb, Sabarish V Babu, Wen-Chieh Lin, and Yun-Xuan Lin. Comparative evaluation of the effects of motion control on cybersickness in immersive virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 672–681. IEEE, 2020.

[218] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Chih-Han Chung, Sabarish V Babu, and Yu-Shuen Wang. Investigating a combination of input modalities, canvas geometries, and inking triggers on on-air handwriting in virtual reality. *ACM Transactions on Applied Perceptions (TAP)*, 2022.

[219] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Balagopal Raveendranath, Christopher C Pagano, Andrew C Robb, Wen-Chieh Lin, and Sabarish 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.

[220] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Balagopal Raveendranath, Christopher C Pagano, Andrew C Robb, Wen-Chieh Lin, and Sabarish 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.

[221] Matias Volonte, Reza Ghaiumy Anaraky, Rohith Venkatakrishnan, Roshan Venkatakrishnan, Bart P Knijnenburg, Andrew T Duchowski, and Sabarish V Babu. Empirical evaluation and pathway modeling of visual attention to virtual humans in an appearance fidelity continuum. *Journal on Multimodal User Interfaces*, 15(2):109–119, 2021.

[222] Tijana Vuletic, Alex Duffy, Laura Hay, Chris McTeague, Gerard Campbell, and Madeleine Grealy. Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, 129:74–94, 2019.

[223] Jaye Wald and Steven Taylor. Efficacy of virtual reality exposure therapy to treat driving phobia: a case report. *Journal of behavior therapy and experimental psychiatry*, 31(3-4):249–257, 2000.

[224] Kai Wang, Daisuke Iwai, and Kosuke Sato. Supporting trembling hand typing using optical see-through mixed reality. *IEEE Access*, 5:10700–10708, 2017.

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

[226] William H Warren. Perceiving affordances: visual guidance of stair climbing. *Journal of experimental psychology: Human perception and performance*, 10(5):683, 1984.

[227] William H Warren Jr and Suzanne Whang. Visual guidance of walking through apertures: body-scaled information for affordances. *Journal of experimental psychology: human perception and performance*, 13(3):371, 1987.

[228] 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`.

[229] Peter J Werkhoven and Joris Groen. Manipulation performance in interactive virtual environments. *Human Factors*, 40(3):432–442, 1998.

[230] Matt Whitlock, Ethan Harnner, Jed R Brubaker, Shaun Kane, and Danielle Albers Szafir. Interacting with distant objects in augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 41–48. IEEE, 2018.

[231] Jason Wither and Tobias Hollerer. Evaluating techniques for interaction at a distance. In *Eighth International Symposium on Wearable Computers*, volume 1, pages 124–127. IEEE, 2004.

[232] Jessica K Witt, Dennis R Proffitt, and William Epstein. Tool use affects perceived distance, but only when you intend to use it. *Journal of experimental psychology: Human perception and performance*, 31(5):880, 2005.

[233] Erik Wolf, Marie Luisa Fiedler, Nina Döllinger, Carolin Wienrich, and Marc Erich 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)*, pages 350–359. IEEE, 2022.

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

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

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

[237] Feng Zhou, Henry Been-Lirn Duh, and Mark 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*, pages 193–202. IEEE, 2008.