How Virtual Hand Representations Affect the Perceptions of Dynamic Affordances in Virtual Reality

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

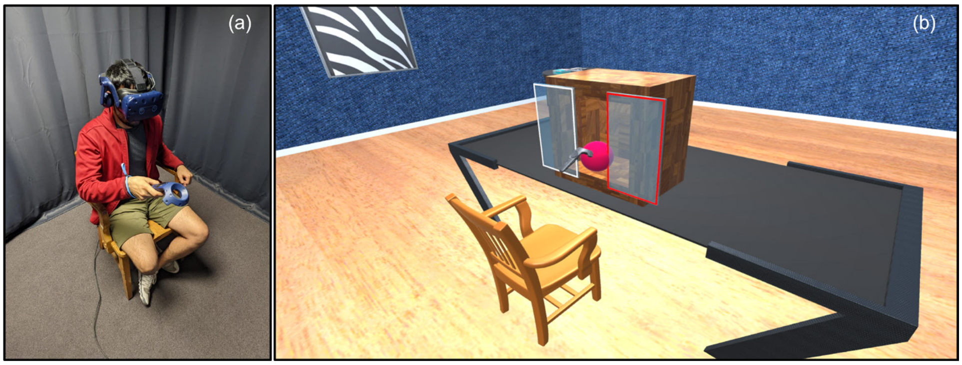


Fig. 1: Participant performing the task from a third person perspective in (a) the real world, and (b) the virtual world. The highlighted door in (b) denotes a failed trial wherein the participant failed to successfully retrieve the object without collisions.

Abstract—User representations are critical to the virtual experience, and involve both the input device used to support interactions as well as how the user is virtually represented in the scene. Inspired by previous work that has shown effects of user representations on the perceptions of relatively static affordances, we attempt to investigate how end-effector representations affect the perceptions of affordances that dynamically change over time. Towards this end, we empirically evaluated how different virtual hand representations affect users' perceptions of dynamic affordances in an object retrieval task wherein users were tasked with retrieving a target 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). Results indicated that the controller-hand condition produced lower levels of performance than both the other conditions. Furthermore, users in this condition exhibited a diminished ability to calibrate their performance over trials. Overall, we find that representing the end-effector as a hand tends to increase embodiment but can also come at the cost of performance, or an increased workload due to a discordant mapping between the virtual representation and the input modality used. It follows that VR system designers should carefully consider the priorities and target requirements of the application being developed when choosing the type of end-effector representation for users to embody in immersive virtual experiences.

Index Terms—Affordance, Passability, Self-Avatar, Virtual Reality

1 Introduction

The rapid surge of virtual reality technology is evident with its applications finding their place in the areas of training [28], education [14], therapy [81], and sports [42], to name a few. The success of these applications has largely been underpinned by advancements in user interactions, an essential component of immersive virtual environments (IVEs). In VR, users typically interact with virtual integrants like objects, humans, information, etc., that convolve together to form the overall experience. There are a plethora of ways to design the interaction between the user and a virtual entity. Grasping an object for instance, can be designed such that the grasping entity either disappears, sticks, fades, highlights, or snaps the object being interacted with to it. Interactions strongly influence users' perceptions of the virtual experience just as how they are central to our experience in the real world, making their mechanics noteworthy aspects that are highly germane to the overall experience.

In VR, interactions can be facilitated through direct VR handheld controller inputs (HTC Vive, Oculus touch, etc.), camera vision based tracked hand gestures (Oculus quest 2, Valve index, Leap motion controller, etc.) or even through supplemental hardware (tracked data gloves using reflective markers) that have been configured to work with the system. These interaction modalities differ in characteristics like fidelity, latency, tracking error(s), etc., each affecting how effectively, efficiently, and accurately users perform actions in the virtual world. Researchers hence continue to investigate how different interaction metaphors affect users' abilities to perform tasks in virtual experiences, attempting to outline scenarios in which an interaction modality aptly

- *Roshan and Rohith Venkatakrishnan are PhD candidates in the School of Computing at Clemson University. E-mail: rvenkat@g.clemson.edu, rohithv@g.clemson.edu*
- *Balagopal Raveendranath is a PhD student with the Department of Psychology at Clemson University. E-mail: braveen@g.clemson.edu*
- *Christopher C. Pagano is a faculty in the Department of Psychology at Clemson University. E-mail: cpagano@clemson.edu*
- *Wen-Chieh Lin is a faculty in the Department of Computer Science at National Yang Ming Chiao Tung University. E-mail: wclin@cs.nctu.edu.tw*
- *Andrew C. Robb and Sabarish V. Babu are faculty in the School of Computing at Clemson University. E-mail: arobb@clemson.edu, sbabu@clemson.edu*

Manuscript received 14 October 2022; revised 13 January 2023; accepted 30 January 2023. Date of publication 22 February 2023; date of current version 29 March 2023. Digital Object Identifier no. 10.1109/TVCG.2023.3247041

complements the performance of some task.

To facilitate interaction in the virtual world, users are usually provided with virtual representations called user representations [71]. These representations provide users with a frame of reference of themselves, visually depicting the locus of the user in the virtual scene. User representations can range from full-body avatars to more visually primitive approaches wherein virtual models of the controllers alone are rendered. The nature and fidelity of these user representations affect users' sense of embodiment, ownership, and agency all of which tend to be important aspects of such representations in interactive IVEs [5]. Given the challenges and additional hardware requirements involved in provisioning tracked full-body avatars, it is common to only represent the user's end-effectors (the entities that interact with the world). In such situations, the user only sees an assigned representation of the virtual controller and acts based on visual, tactile, and proprioceptive information afforded by this visualization of the end-effector. In doing so, users incorporate the assigned representation into their body schema which has been shown to be fluid and malleable [17]. In attempting to maximize embodiment and ownership, contemporary VR experiences are resorting to representing the end-effectors as hands as opposed to virtual replicas of the handheld controllers themselves. This leads to a discordant mapping between what users see and use as their end-effectors, potentially affecting interactions. End-effector representations can also vary anthropometrically or anthropomorphically, each having its own consequences on how IVEs and affordances within the experience are perceived [19, 57].

Affordances are the relationship between the environment and the organism/actor, and determine one's opportunities for action like passability, graspability, reachability, etc., [8, 24]. They can be static or can even exhibit dynamic properties wherein elements of the environment temporally move, requiring the actor to synchronize their own movements to the movement of the objects in the environment. Examples of such dynamic affordances include crossing gaps in moving traffic, stepping on a moving escalator, catching a moving ball, etc. Affordances scale to the organism, and are determined by the morphology and physical capabilities of the actor [24]. For example, objects are judged smaller when one's virtual hand is larger [57]. Researchers continue to show that the way in which a user and their end-effectors are represented in VR, affect their perceptions and action capabilities in these virtual experiences.

The appearance of users' self avatars have been demonstrated to have effects on whether or not apertures are judged as passable [64]. While such work has been extensively researched in contexts involving relatively immutable affordances, limited work delves into more dynamic contexts. Research on dynamic affordances conducted in IVEs has explored the effects of display type, method of locomotion, and age on gap crossing behavior, finding significant impacts of these factors [15, 25]. However, little is known about the effects of user representations on user behavior, and it remains to be seen as to how the virtual end-effector representation affects task performance in complex interactions that encompass dynamic affordances. Given the growing number of VR training, rehabilitation and sport related applications featuring near-field interactions with rather dynamic elements [38, 50], it is imperative for researchers to understand how design decisions made on end-effector representations affect these interactions in the VR experience. Apropos of this, we discuss the results of an experimental investigation that evaluates how end-effector representations affect near field interactions with dynamic affordances in IVEs. Downstream of this, 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.

2 Related Works

2.1 Grasping in VR

Interactions with virtual objects is a topic that has been extensively explored [4, 26, 31, 69]. Grasping is a common interaction that people perform routinely and its realistic simulation in VR requires dedicated hardware and algorithms [9]. It can be treated as an egocentric manipulation technique when simulated using virtual hands which in turn allows users to hold virtual objects upon contact with the hands [66]. While visualizing grasping, depicting the grasping entity 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, 12]. Interaction metaphors that involve virtual hands are improved by increasing users' control over the hands by affording finger motions, and by providing visual feedback [37, 76]. Furthermore, it bodes well to provide additional feedback through the interacting entities (virtual hands and objects) by using illumination effects [76] and indicating the grasping status [53]. Research suggests that high fidelity grasping with tracked gloves using IMU sensors performs better than camera vision based commodity hardware like the LeapMotion sensor [43].

Fig. 2: The input devices used in the experiment. a)HTC Vive Pro controller, and b)Noitom's Hi5 VR glove

2.2 User representations: Interaction and Embodiment

In IVEs, the user is often provided with virtual representations that are commonly referred to as user representations [71, 72]. These representations help users perform actions and affect how effectively, accurately, and efficiently users can perform them [19, 52, 54, 77]. Furthermore, they play a role in shaping users' sense of embodiment, a phenomenon that comprises the sense of self-location, agency, and ownership [35,45], which are important characteristics associated with interactive IVEs. De Vignemont [18] 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 [7], while ownership pertains to one's self attribution of a body [23, 80]. 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 [72].

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 [49], Noitom's Hi5 VR gloves with 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 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 [11]. The visual and kinematic properties of a virtual hand representation and its corresponding motor synchronicity can also affect task performance [13, 60]. Other studies have shown that virtual hands do not significantly impact performance for tasks involving tool based object manipulation [44, 67].

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 [40]. 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 [51, 69]. 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 [2]. 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 [34]. Overall, 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 controllers producing better performance [53, 74].

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 [5, 35, 36]. 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 [63]. Research further suggests that users experience lower cognitive loads while performing a spatial rotation task in the presence of self avatars than without [77]. 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 [54]. 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 [19]. 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 [41, 57]. 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 [58].

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 [3]. 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 [46]. 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 [5]. 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 [79]. 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 [48]. 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 Affordances and User representations

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 [8]. Stated simply, they refer to the opportunities of actions offered by the surrounding environment to the perceiver. 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 [30]. For example, the shoulder width of an individual determines their ability to frontally pass through an aperture [82, 83]. 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 [21, 68]. It hence follows that manipulations of an individual's body schema or their action capabilities affects the way they perceive the environment [78, 83]. This has also been demonstrated in contexts of virtual environments wherein manipulations of the user's self avatar or their action capability has been shown to affect perceptions of depth, size, weight, and passability [19, 57].

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 judgements on whether gaps can be perceived as crossable [33]. Similarly, the appearance of users' self avatars (i.e, overweight/underweight) was shown to affect passability judgements for apertures created between two poles [64], demonstrating that user representations affect how affordances are perceived [6]. In these studies, the affordances of the gaps and apertures are static in that they exist in the environment with unaltered spatio-temporal dynamics. However, interactions can also involve affordances that exhibit dynamic properties. Examples 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 [10, 16, 22, 47], and crossing streets with oncoming traffic [15, 59, 65]. In IVEs, such research has shown that the locomotion method (action capability) affects the perceived opportunities for action in the environment [25]. 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 judgements 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 [15, 65, 73]. 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 perceptions of static affordances in IVEs [20, 39, 54], there is merit in exploring if and how they affect perception of affordances in more dynamic contexts, and this work seeks to contribute to that knowledge base.

3 System Description

3.1 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 90 Hz. The HTC Vive Pro controller and the Noitom Hi5 VR glove (figure 2) 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 Virtual Environment

A 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 to provide a realistic sense of size and scale with familiar objects in the experiment environment. A virtual wooden box was placed on a virtual table in the center of the room and this box was used to host the near-field object retrieval task described in section 4.1. 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 1). The virtual wooden chair was modeled and textured to match the dimensions and look of its real world counterpart.

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 with specular reflection 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 alone. 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.

3.3 End-Effector Representations

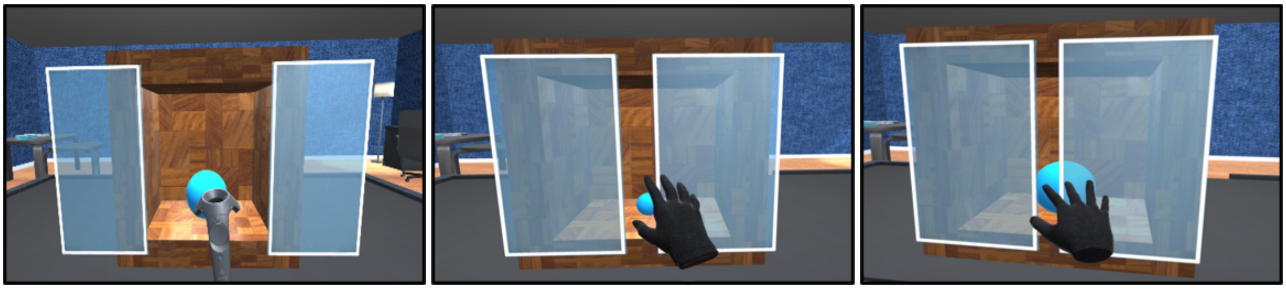
This study investigated three different virtual end-effector representations the specifics of which are described in this section. We use the term end-effector as it encompasses terms that refer to either the controller, the virtual hand or more broadly speaking, the tool with which users interact in 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 users' 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 based on the button being pressed, accounting for 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 [56]. 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 [46]. 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. An adult (large) sized right-handed glove was procured for this condition. Users wearing this glove could grab the target ball by simply clenching their fists. The prehension 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 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 [70].

A system evaluation of latency in all three conditions was conducted using Niehorster et al.'s method [55]. Ten samples of latency 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 the physical device (controller or a hand when wearing the glove) as well as their respective end-effector representations through a monocular view port of the HMD (with the lens removed). The device was moved in a straight line (translational), and



(a) Controller (b) Controller-hand (c) Glove

Fig. 3: First person 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 figure (a) shows the doors sliding out towards the widest open aperture width, while (b) and (c) show the doors sliding in towards the close position.

rotated about the vertical axis (rotational) multiple times, thus capturing several respective latency samples. Using video editing software, the translational and rotational latency were calculated for each trial. The mean end-to-end latency of the conditions were as follows: Controller (Pos. lag =14.16ms and Ori. lag = 14.58ms), Controller-hand (Pos. lag = 14.58ms, Ori. lag=15ms), Glove (Pos.lag = 11.25ms, Ori. lag = 10.41ms).

4 Experiment

4.1 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 box was situated on virtual table in front of the user (figures 3 and 1). The glass doors in front allowed users to see the target during the trials. Users were seated on a wooden chair that was physically co-located with a virtual 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, 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 successfully retrieve the ball in each trial.

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 (or both) 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 1), 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. Upon success or failure, users had to move their hand to the start position by placing their end-effector on the red bubble on the arm rest to initiate the next trial. 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 and reset to the close position. It was ensured that the success and collision sounds were 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.

4.2 Study Design

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. Users in VR often interact with targets of different sizes, which in turn can influence how difficult it is to perform such a task. In comparison to a small target, a large target may be easier to reach but harder to manipulate through an aperture while avoiding collisions. This merits the investigation of target size on performance. In addition to making the task more like real world reaching, the variation in ball size introduces additional variation in motor control that will require the participants to maintain a high level of attention to the task.

For each experimental condition, participants performed the object retrieval task described in section 4.1 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.

4.3 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 an estimate of the probability of successfully retrieving the target 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. [62]. This questionnaire comprises 16 items that load on to four interrelated sub-dimensions, 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 [27].

4.4 Research Question and Hypotheses

This study aimed at 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 how such representations of the virtual end-effector affect perceived levels of embodiment and workload. User performance was operationalized based on the measure described in section 4.3. The following hypotheses were developed:
H1: The controller condition will produce the best performance.
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.

4.5 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 which whom identified as female. All participants had normal or corrected-to-normal (20/20) vision, along with normal motor function of their upper body.

4.6 Procedure

Upon arrival to the laboratory, participants were greeted and asked to read and sign a consent form. 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 4.1). 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, and 150 oscillations per minute (Hz). 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 each participant in the glove condition, a calibration procedure was performed right before the practice phase to precisely calibrate the representation, its finger pose tracking, position, and orientation to precisely match the user's actual hand. 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, and the NASA TLX questionnaire. They then participated in a debriefing interview about the study and were compensated for their time. On average, it took a participant up to 50 minutes to complete the whole procedure.

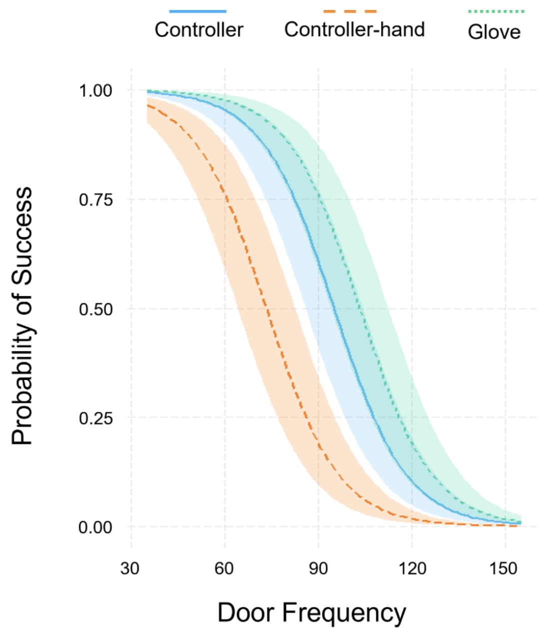


Fig. 4: Effect of door frequency on success, moderated by condition. Shading around each line indicates 95% confidence intervals.

5 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 [29]. 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 [75].

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

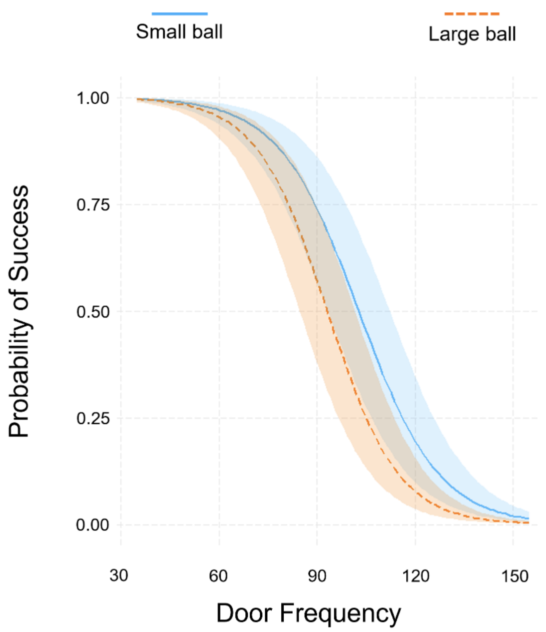


Fig. 5: Effect of door frequency on success rate, moderated by ball size. Shading around each line indicates 95% confidence intervals

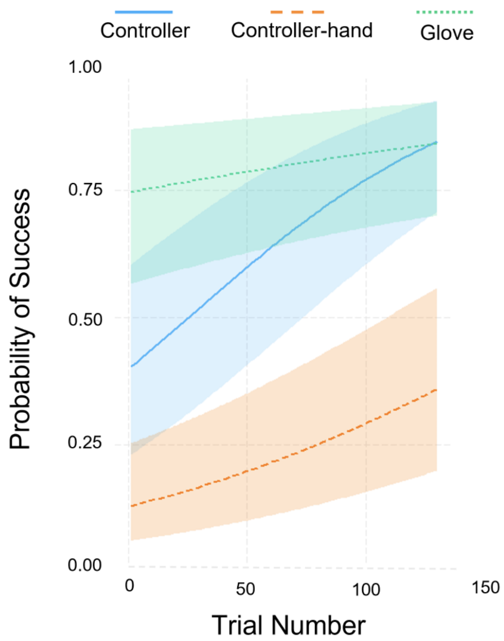


Fig. 6: Effect of trial number on success rate, moderated by condition. Shading around each line indicates 95% confidence intervals.

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 $R^2 = 0.655$).

5.1 Performance

5.1.1 End-Effector Representation

As expected, there was a significant effect of condition on success rate, $\chi^2(2, N = 7800) = 23.02$, $p < 0.001$, $sr^2 = 0.067$. 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.

5.1.2 Frequency, Target Size and Learning

There was a significant effect of door frequency on success rate, $\chi^2(1, N = 7800) = 1866.32$, $p < 0.001$, $sr^2 = 0.60$. 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.

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 was smaller (M probability = 0.56, SE = 0.06) than when larger (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.

5.1.3 Interaction Effects

Condition was a significant moderator for the effect of door frequency, $\chi^2(2, N = 7800) = 60.75$, $p < 0.001$, $sr^2 = 0.02$. The levels of condition altered the relationship between door frequency and success rate. As seen in figure 4, 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, $p < 0.001$). Although the success rate generally reduced with increasing door frequency, it reduced at a much slower rate in the glove condition (indicating a superior performance in this condition) as compared to the other conditions.

Ball size was also a significant moderator for the effect of door frequency, $\chi^2(1, N = 7800) = 13.09$, $p < 0.001$, $sr^2 = 0.002$. As seen in figure 5, 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, $\chi^2(2, N = 7800) = 20.89$, $p < 0.001$, $sr^2 = 0.002$. As seen in figure 6, a test of simple slopes revealed that for each condition, the simple slope for trial number was positive, suggesting a learning effect from one trial to the next. 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, $\chi^2(3, N = 7800) = 12.69$, $p = 0.005$, $sr^2 = 0.002$. As seen in figure 7, a test of simple slopes for trial number revealed that, when the door frequency was 57.58 (-1 standard deviation), 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), 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),

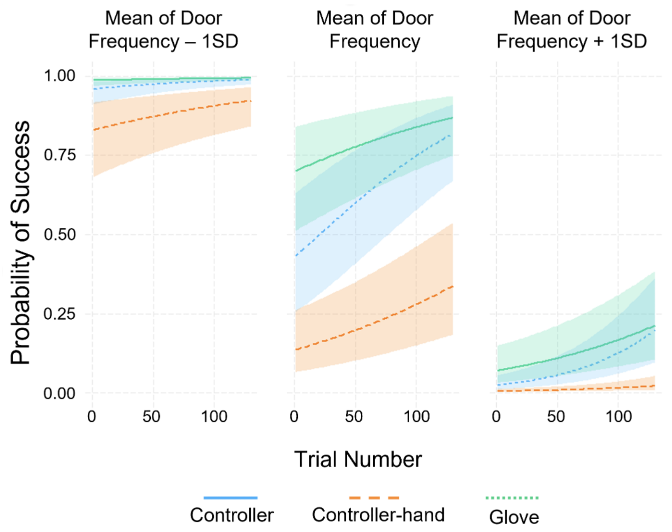


Fig. 7: Performance over trials in the conditions at (a) one standard deviation below the mean door frequency, (b) mean door frequency, and (c) one standard deviation above the mean door frequency. Shading around each line indicates 95% confidence intervals.

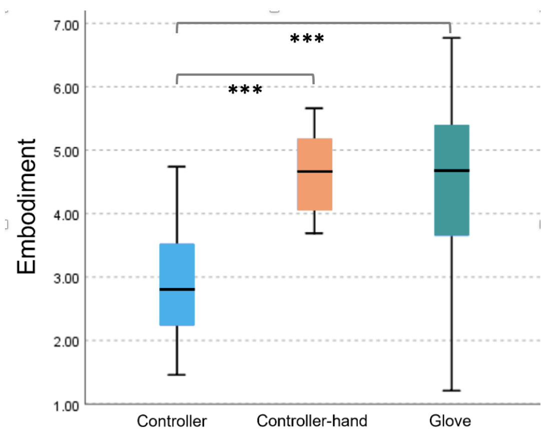


Fig. 8: Embodiment scores across the different conditions.

the slope of trial number was again significantly different from zero for the 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, $p < 0.001$).

5.2 Embodiment

A Kruskal-Wallis H test showed that there was a statistically significant difference in the perceived embodiment score between the different conditions, $\chi^2(2) = 25.51$, $p < 0.001$. 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 embodiment was found between the controller and controller-hand condition $U(N_{controller} = 20, N_{controller-hand} = 20) = 26.0$, $z = -4.7$, $p < 0.001$, and between the controller and glove condition $U(N_{controller} = 20, N_{glove} = 20) = 53.0$, $z = -3.97$, $p < 0.001$. Participants in the controller condition (Mdn = 2.81) perceived less overall embodiment than those in the controller-hand (Mdn = 4.66), and glove (Mdn = 4.68) conditions. See figure 8.

5.3 Workload

A Kruskal-Wallis H test showed that there was a statistically significant difference in the perceived workload score between the different conditions, $\chi^2(2) = 8.627$, $p < 0.013$. 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 workload was found between the controller and controller-hand conditions $U(N_{controller} = 20, N_{controller-hand} = 20) = 106.0$, $z = -2.54$, $p = 0.011$, and between the controller-hand and glove conditions $U(N_{controller} = 20, N_{glove} = 20) = 109.5$, $z = -2.45$, $p = 0.014$. Participants in the controller-hand condition (Mdn = 7.1) perceived more overall workload than participants in the controller (Mdn = 5.7), and glove (Mdn = 6.2) conditions. See figure 9.

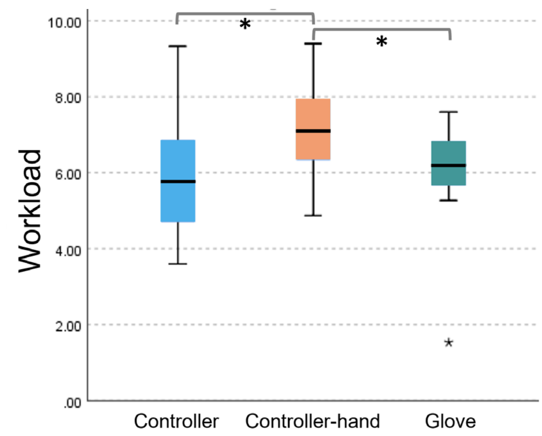


Fig. 9: Workload scores in the different conditions of the study.

6 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 major differences between the controller and glove conditions, but that the controller-hand condition negatively impacted performance. Figure 4 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 [46]. Furthermore, our findings situate with research showing that the end-effector representation affects the perception of available action possibilities in virtual reality [32], 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 [61] 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 4 and 5, 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 5). 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 6. 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 [32]. 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, 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 [46]. 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. Overall, our work serves to show that the end-effector representation and its mapping with the input modality used strongly influences how well users are able to perform tasks that involve dynamic affordances and to what extent they embody their virtual representations.

7 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 [47]. 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.

8 Conclusion and Future Work

In this work, we empirically evaluated how different virtual end-effector representations affect users' perceptions of dynamic affordances in an object retrieval task in the near field. Users performed a collision-avoidance based object retrieval task for a number of trials. We employed a between-subjects study design manipulating the end-effector representation across three experimental conditions, namely, Controller (using a controller represented as a virtual controller), Controller-hand (using a controller represented as a virtual hand), and Glove (using a hand tracked hi-fidelity glove represented as a virtual hand). Results indicated 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 over trials. Overall, we find that representing the end-effector as a hand can be detrimental to performance when there is a discordant mapping between the virtual representation and the input modality used.

In this line of work, we wish to further investigate how our 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 hooks and prosthetic limbs to photorealistic 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 interest. We call for future investigations to further explore what other tradeoffs can manifest as a result of discordant mappings between user representations and input modalities used to support interactions.

Acknowledgments

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

References

[1] A. Adkins, L. Lin, A. Normoyle, R. Canales, Y. Ye, and S. Jörg. Evaluating grasping visualizations and control modes in a vr game. *ACM Transactions on Applied Perception (TAP)*, 18(4):1–14, 2021. 2, 3

[2] A. Alzayat, M. Hancock, and M. 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*, pp. 1–11, 2019. 3

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

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

[5] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE virtual reality (VR)*, pp. 3–10. IEEE, 2016. 2, 3

[6] A. Bhargava, R. Venkatakrishnan, R. Venkatakrishnan, H. Solini, K. M. Lucaites, A. Robb, C. Pagano, and S. 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*, 2021. 3

[7] O. Blanke and T. Metzinger. Full-body illusions and minimal phenomenal selfhood. *Trends in cognitive sciences*, 13(1):7–13, 2009. 2

[8] J. J. Blau and J. B. Wagman. *Introduction to Ecological Psychology: A lawful approach to perceiving, acting, and cognizing*. Taylor & Francis, 2022. 2, 3

[9] C. W. Borst and A. P. Indugula. Realistic virtual grasping. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 91–98. IEEE, 2005. 2

[10] M. Buekers, G. Montagne, A. de Rugy, and M. Laurent. The regulation of externally paced human locomotion in virtual reality. *Neuroscience Letters*, 275(3):171–174, 1999. 4

[11] G. Caggianese, L. Gallo, and P. 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*, pp. 24–33. Springer, 2018. 2

[12] R. Canales, A. Normoyle, Y. Sun, Y. Ye, M. D. Luca, and S. Jörg. Virtual grasping feedback and virtual hand ownership. In *ACM Symposium on Applied Perception 2019*, pp. 1–9, 2019. 2

[13] J. C. Castro-Alonso, P. Ayres, and F. Paas. Dynamic visualisations and motor skills. In *Handbook of human centric visualization*, pp. 551–580. Springer, 2014. 3

[14] Y.-T. Chen, C.-H. Hsu, C.-H. Chung, Y.-S. Wang, and S. 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)*, pp. 172–180. IEEE, 2019. 1

[15] B. J. Chihak, J. M. Plumert, C. J. Ziemer, S. Babu, T. Grechkin, J. F. Cremer, and J. 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. 2, 4

[16] M. E. Cinelli, A. E. Patla, and F. Allard. Behaviour and gaze analyses during a goal-directed locomotor task. *Quarterly Journal of Experimental Psychology*, 62(3):483–499, 2009. 4

[17] B. Day, E. Ebrahimi, L. S. Hartman, C. C. Pagano, A. C. Robb, and S. V. Babu. Examining the effects of altered avatars on perception-action in virtual reality. *Journal of Experimental Psychology: Applied*, 25(1):1, 2019. 2

[18] F. De Vignemont. Embodiment, ownership and disownership. *Consciousness and cognition*, 20(1):82–93, 2011. 2

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

[20] E. Ebrahimi, A. Robb, L. S. Hartman, C. C. Pagano, and S. V. Babu. Effects of anthropomorphic fidelity of self-avatars on reach boundary estimation in immersive virtual environments. In *Proceedings of the 15th ACM Symposium on Applied Perception*, pp. 1–8, 2018. 4

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

[22] B. R. Fajen and J. S. Matthis. Direct perception of action-scaled affordances: the shrinking gap problem. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5):1442, 2011. 4

[23] S. Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21, 2000. 2

[24] J. J. Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014. 2

[25] T. Y. Grechkin, J. M. Plumert, and J. 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. 2, 4

[26] K. S. Hale and K. M. Stanney. *Handbook of virtual environments: Design, implementation, and applications*. CRC Press, 2014. 2

[27] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988. 6

[28] W. L. Heinrichs, P. Youngblood, P. M. Harter, and P. Dev. Simulation for team training and assessment: case studies of online training with virtual worlds. *World journal of surgery*, 32(2):161–170, 2008. 1

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

[30] S. Ishak, K. E. Adolph, and G. C. Lin. Perceiving affordances for fitting through apertures. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6):1501, 2008. 3

[31] J. Jankowski and M. Hachet. A survey of interaction techniques for interactive 3d environments. In *Eurographics 2013-STAR*, 2013. 2

[32] T. Joy, E. Ugur, and I. 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. 8, 9

[33] E. Jun, J. K. Stefanucci, S. H. Creem-Regehr, M. N. Geuss, and W. 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. 3

[34] C. Khundam, V. Vorachart, P. Preeyawongsakul, W. Hosap, and F. Noël. A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. In *Informatics*, vol. 8, p. 60. Multidisciplinary Digital Publishing Institute, 2021. 3

[35] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012. 2, 3

[36] E. Kokkinara and M. 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. 3

[37] P. G. Kry, A. Pihuit, A. Bernhardt, and M.-P. Cani. Handnavigator: Hands-on interaction for desktop virtual reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pp. 53–60, 2008. 2

[38] K. E. Laver, B. Lange, S. George, J. E. Deutsch, G. Saposnik, and M. Crotty. Virtual reality for stroke rehabilitation. *Cochrane database of systematic reviews*, (11), 2017. 2

[39] M. Leyrer, S. A. Linkenauger, H. H. Bülthoff, U. Kloos, and B. 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*, pp. 67–74, 2011. 4

[40] L. Lin, A. Normoyle, A. Adkins, Y. Sun, A. Robb, Y. Ye, M. Di Luca, and S. 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)*, pp. 510–518. IEEE, 2019. 3

[41] S. A. Linkenauger, M. Leyrer, H. H. Bülthoff, and B. 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. 3

[42] H. Liu, Z. Wang, C. Mousas, and D. Kao. Virtual reality racket sports: Virtual drills for exercise and training. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 566–576. IEEE, 2020. 1

[43] H. Liu, Z. Zhang, X. Xie, Y. Zhu, Y. Liu, Y. Wang, and S.-C. Zhu. High-fidelity grasping in virtual reality using a glove-based system. In *2019 international conference on robotics and automation (icra)*, pp. 5180–5186. IEEE, 2019. 2

[44] B. Lok, S. Naik, M. Whitton, and F. 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. 3

[45] M. R. Longo, F. Schüür, M. P. Kammers, M. Tsakiris, and P. Haggard.

What is embodiment? a psychometric approach. *Cognition*, 107(3):978–998, 2008. 2

[46] C. Lougiakis, A. Katifori, M. Roussou, and I.-P. Ioannidis. Effects of virtual hand representation on interaction and embodiment in hmd-based virtual environments using controllers. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 510–518. IEEE, 2020. 3, 4, 8, 9

[47] K. M. Lucaites, R. Venkatakrishnan, R. Venkatakrishnan, A. Bhargava, and C. C. Pagano. Predictability and variability of a dynamic environment impact affordance judgments. *Ecological Psychology*, 32(2-3):95–114, 2020. 4, 9

[48] J.-L. Lugrin, M. Ertl, P. Krop, R. Klüpfel, S. Stierstorfer, B. Weisz, M. Rück, J. Schmitt, N. Schmidt, and M. E. Latoschik. Any "body" there? avatar visibility effects in a virtual reality game. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 17–24. IEEE, 2018. 3

[49] Manus. Wearable Manus Gloves, 2022. https://www.manus-meta.com/products/gloves. 2

[50] E. Matsas and G.-C. Vosniakos. Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 11(2):139–153, 2017. 2

[51] R. P. McMahan, A. J. D. Alon, S. Lazem, R. J. Beaton, D. Machaj, M. Schaefer, M. G. Silva, A. Leal, R. Hagan, and D. A. Bowman. Evaluating natural interaction techniques in video games. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 11–14. IEEE, 2010. 3

[52] E. A. McManus, B. Bodenheimer, S. Streuber, S. De La Rosa, H. H. Bülthoff, and B. J. Mohler. The influence of avatar (self and character) animations on distance estimation, object interaction and locomotion in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on applied perception in graphics and visualization*, pp. 37–44, 2011. 2

[53] M. Moehring and B. Froehlich. Effective manipulation of virtual objects within arm's reach. In *2011 IEEE Virtual Reality Conference*, pp. 131–138. IEEE, 2011. 2, 3

[54] B. J. Mohler, S. H. Creem-Regehr, W. B. Thompson, and H. H. Bülthoff. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence*, 19(3):230–242, 2010. 2, 3, 4

[55] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017. 4

[56] Noitom. Noitom Hi5 Glove, 2022. https://www.noitom.com/hi5-vr-glove. 4

[57] N. Ogawa, T. Narumi, and M. Hirose. Virtual hand realism affects object size perception in body-based scaling. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 519–528. IEEE, 2019. 2, 3

[58] N. Ogawa, T. Narumi, H. Kuzuoka, and M. 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*, pp. 1–14, 2020. 3

[59] E. E. O'Neal, Y. Jiang, L. J. Franzen, P. Rahimian, J. P. Yon, J. K. Kearney, and J. 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. 4

[60] O. Ossmy and R. Mukamel. Short term motor-skill acquisition improves with size of self-controlled virtual hands. *PloS one*, 12(1):e0168520, 2017. 3

[61] C. C. Pagano and B. Day. Ecological interface design inspired by "the meaningful environment". In *Perception as Information Detection*, pp. 37–50. Routledge, 2019. 9

[62] T. C. Peck and M. Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 1:575943, 2021. 6

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

[64] I. V. Piryankova, H. Y. Wong, S. A. Linkenauger, C. Stinson, M. R. Longo, H. H. Bülthoff, and B. J. Mohler. Owning an overweight or underweight body: distinguishing the physical, experienced and virtual body. *PloS one*, 9(8):e103428, 2014. 2, 3

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

[66] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum*, vol. 17, pp. 41–52. Wiley Online Library, 1998. 2

[67] A. Ricca, A. Chellali, and S. 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)*, pp. 260–268. IEEE, 2020. 3

[68] S. Runeson and G. 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. 3

[69] A. Schäfer, G. Reis, and D. Stricker. The gesture authoring space: Authoring customised hand gestures for grasping virtual objects in immersive virtual environments. *arXiv e-prints*, pp. arXiv–2207, 2022. 2, 3

[70] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. " these are not my hands!" effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1577–1582, 2017. 4

[71] S. Seinfeld, T. Feuchtner, A. Maselli, and J. Müller. User representations in human-computer interaction. *Human–Computer Interaction*, 36(5-6):400–438, 2021. 2

[72] S. Seinfeld, T. Feuchtner, J. Pinzek, and J. Müller. Impact of information placement and user representations in vr on performance and embodiment. *IEEE transactions on visualization and computer graphics*, pp. 1–13, 2020. 2

[73] G. Simpson, L. Johnston, and M. Richardson. An investigation of road crossing in a virtual environment. *Accident Analysis & Prevention*, 35(5):787–796, 2003. 4

[74] P. Skalski, R. Tamborini, A. Shelton, M. Buncher, and P. Lindmark. Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. *New Media & Society*, 13(2):224–242, 2011. 3

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

[76] J. Sreng, A. Lécuyer, C. Mégard, and C. 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. 2

[77] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE virtual reality (VR)*, pp. 67–76. IEEE, 2016. 2, 3

[78] J. K. Stefanucci and M. N. Geuss. Big people, little world: The body influences size perception. *Perception*, 38(12):1782–1795, 2009. 3

[79] T. Q. Tran, H. Shin, W. Stuerzlinger, and J. Han. Effects of virtual arm representations on interaction in virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–9, 2017. 3

[80] M. Tsakiris, G. Prabhu, and P. Haggard. Having a body versus moving your body: How agency structures body-ownership. *Consciousness and cognition*, 15(2):423–432, 2006. 2

[81] J. Wald and S. 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. 1

[82] W. H. Warren. Perceiving affordances: visual guidance of stair climbing. *Journal of experimental psychology: Human perception and performance*, 10(5):683, 1984. 3

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