Hand Interfaces: Using Hands to Imitate Objects in AR/VR for Expressive Interactions

Siyou Pei
University of California, Los Angeles
Los Angeles, CA, USA
sypei@g.ucla.edu

Alexander Chen
University of California, Los Angeles
Los Angeles, CA, USA
yoyohc10@g.ucla.edu

Jaewook Lee
University of Illinois at Urbana-Champaign
Urbana, IL, USA
jaewook4@illinois.edu

Yang Zhang
University of California, Los Angeles
Los Angeles, CA, USA
yangzhang@ucla.edu

Figure 1: Hand Interfaces allow users to imitate a wide range of objects that we perceive as AR/VR interfaces for expressive, readily available interactions.

ABSTRACT

Augmented reality (AR) and virtual reality (VR) technologies create exciting new opportunities for people to interact with computing resources and information. Less exciting is the need for holding hand controllers, which limits applications that demand expressive, readily available interactions. Prior research investigated freehand AR/VR input by transforming the user's body into an interaction medium. In contrast to previous work that has users' hands grasp virtual objects, we propose a new interaction technique that lets users' hands become virtual objects by imitating the objects themselves. For example, a thumbs-up hand pose is used to mimic a joystick. We created a wide array of interaction designs around this idea to demonstrate its applicability in object retrieval and interactive control tasks. Collectively, we call these interaction designs Hand Interfaces. From a series of user studies comparing

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3501898

Hand Interfaces against various baseline techniques, we collected quantitative and qualitative feedback, which indicates that Hand Interfaces are effective, expressive, and fun to use.

CCS CONCEPTS

• **Human-centered computing → Interaction techniques**.

KEYWORDS

AR/VR, Interaction design, Imitation, Embodiment, Free-hand interactions, On-body interactions

ACM Reference Format:
Siyou Pei, Alexander Chen, Jaewook Lee, and Yang Zhang. 2022. Hand Interfaces: Using Hands to Imitate Objects in AR/VR for Expressive Interactions. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3491102.3501898

1 INTRODUCTION

AR and VR have shown great promise in education [27], accessibility [51], and health care [18]. Paired with a rapidly growing user base, it will likely be the next ubiquitous device (after smartphones) that fundamentally changes human-computer interaction. To support the growing interest towards AR/VR technologies, there has been a significant amount of research on interaction techniques that allow users to easily and naturally manipulate content in the AR/VR space. Conventional input techniques rely on hand-held controllers or in-air gestures. However, one recent research momentum leverages users' hands as an expressive interaction medium. The benefits are multi-fold, many of which come with on-body interactions by default, such as easy and swift access, proprioception, and tactile feedback that allow for more precise control. Aligned with this are prior works that investigated interaction techniques designed around users' hands [5, 15, 26, 38, 58, 65].

Our work is similar in that we also looked into the design space of hand-centric interactions for AR/VR. However, prior works either used hands as 2D surfaces for touch interactions (e.g., *Finger Input* [53], *ActiTouch* [68], *It's a Wrap* [5], *SkinMarks* [55]), or discrete controllers for mode switching (e.g., *Open Palm Menu* [2], Surale et al. [58]). Little has been done considering users' hands as expressive 3D structures to host interactions. One of few prior systems that leveraged the hands' 3D expressivity is *VirtualGrasp* [65], a technique that lets users retrieve virtual objects by performing hand poses as if they were grasping the objects. Hand Interfaces attempts to extend this line of prior work. However, instead of having users' hands ***grasp*** objects, we asked what if we let users' hands ***become*** the objects.

The idea for this research was conceived from the *Rock Paper Scissors* game in which people form hand postures to imitate a rock, a paper, and a pair of scissors. The embodiment (i.e., hands becoming objects) happens in an efficient and self-revealing manner, which contributes to the universality of the game. In fact, we included *Scissors* in Hand Interfaces (Figure 1*q*), and it was among the most well-received designs in our evaluation. Researchers have leveraged embodiment to create interaction techniques that allow users to embody objects with their bodies [9, 10, 22, 52, 59, 66]. This work builds on these previous explorations and dives deeper down into how users' hands can embody a diverse set of objects for two specific tasks in AR/VR – object retrieval and interactive control.

The expressivity of human hands is an innate ability of ours, but an under-explored method of interaction in AR/VR settings. Our objective is to demonstrate the potential of this unique advantage in implementing expressive retrieval and interactivity, and evaluating the pros and cons of this idea with user studies. Specifically, users perform certain hand poses to retrieve corresponding interactive controllers for manipulation, in which users can use the whole hand or a part of it as an input medium. For example, a thumbs-up hand pose imitates a virtual joystick, which the users can control by moving their thumbs around (Figure 1*a*). Additionally, extended fingers can imitate Kalimba keys (Figure 1*j*), while users can play simple music by tapping keys with another hand. In another example, the joint of an index finger alone can emulate a toggle switch (Figure 1*z*). Rest of the examples can be found in Figure 1.

Another significance of this research lies in the evaluation of the benefits of hand-centric user interfaces. In recent years, there has been a surge of research on designing AR/VR interfaces centered around users' hands, but many designs might not have been evaluated with real users (e.g., [41, 42, 62]). This lack of systematic investigation makes it difficult to assess their merits and pitfalls for designers and researchers who want to build upon this line of work. In response, not only did we propose a new interaction technique, we also evaluated it with a wide array of designs (i.e., 11 controllers) in two interaction scenarios (i.e., object retrieval and interactive control) and reported our findings. Additionally, our evaluation included other common techniques including Drop-down Menu, *VirtualGrasp*, Fist Gesture, and Virtual Manipulation as baseline techniques, establishing a foothold for future research to build upon hand-centric user interfaces.

In this research, we first systematically reviewed prior works and guidelines on using bare hands as expressive controls for AR/VR. We then designed a wide array of interaction techniques based on Hand Interfaces and built proof-of-concept detection pipelines with an Oculus Quest headset and its hand tracking feature [40]. Finally, we evaluated Hand Interfaces with 11 distinct interface designs with respect to object retrieval and interactive control in a series of user studies. Both qualitative and quantitative feedback were gathered from 17 participants. Overall, the results indicate that Hand Interfaces are effective, expressive, and fun to use. The advantages of our interaction technique is demonstrated in the three examples below.

Ubiquitous Computing. Hand Interfaces are directly applicable to AR scenarios, in which many applications demand free-hand

Figure 2: A user uses Hand Interfaces in concert with an AR device to quickly and easily control a smart lighting system. Specifically, the *Toggle switch* is used to turn on/off the light, the *Joystick* controls its pan/tilt, and a fist-imitated *Color palette* is used to set its color.

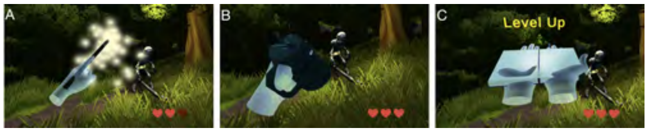

Figure 3: In a magic fighting game, a user uses the hand-imitated *Wand* to cast spells, the *Mug* to drink a healing potion, and the *Book* to level up.

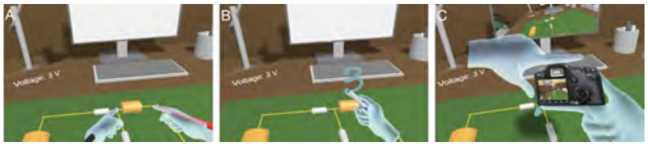

Figure 4: Hand Interfaces allow users to easily use various tools in an educational setting. Specifically, *Multimeter probes* are used to measure voltages, the *Pen* is used for writing notes, and the *Camera* is used for taking a photo of the current scene.

interactions so that users can quickly switch between their tasks in the physical world and the digital world (Figure 2). In this example, a user with a pair of AR glasses can turn a smart lighting system on and off with the *Toggle switch* (C). Then, the user controls the orientation of the light with the *Joystick* (D). Finally, the user retrieves a spherical *Color palette* imitated by a fist gesture to adjust the light color (E).

Entertainment. Hand Interfaces can also be easily applied in VR entertainment applications. In a magic fighting game (Figure 3), players retrieve wands once the *Wand* gesture is performed (A). By waving their wands (index fingers) following specific trajectories, players can strategically cast different spells to win the fight. After taking hits, a player can choose to use a healing potion by performing a *Mug* gesture performing a "drinking" motion with the mug (B). Players can also retrieve other tools; for instance, they can retrieve a *Book*, and level up by opening it (C).

Education. In this VR scenario for circuit education (Figure 4), students can use Hand Interfaces to quickly and easily retrieve tools that facilitate their learning experiences. In this example, a student raises two index fingers to retrieve *Multimeter probes* to measure the voltage of a circuit component (A). Student then uses the *Pen* to write down a note (B). Finally, the student takes a photo of the entire setup using the *Camera* (C).

2 RELATED WORK

In this section, we first talk about design principles and heuristics from prior work on creating effective AR/VR interactions, which were referred to during the creation of Hand Interfaces. Then, we review two key research areas in AR/VR that are closely related to Hand Interfaces.

2.1 AR/VR Interaction Design Principles and Heuristics

Design principles and heuristics help guide interaction technique designs and their evaluations. Such principles and heuristics have been well established for GUIs on conventional computer platforms. One example is the wildly adopted set of heuristics in the evaluation of user interfaces proposed by Nielsen and Molich [46]. As designers and researchers recognize the fundamental differences between 2D and 3D user interfaces, there have been recent efforts in creating design principles and heuristics geared towards AR/VR scenarios. In the consumer domain, interaction designers and developers released posts and blogs to guide developers of their products. For example, Ultraleap posted its guidelines for free-hand AR/VR interactions [62]. Microsoft shared their design philosophy of AR interaction utilizing hands and arms [41, 42]. The increasingly available commercial products of AR/VR have lowered the barrier that has led to a rapidly growing user community of AR/VR, and shared design recommendations online [17, 28, 45, 60].

In the research domain, Dϋnser et al. [12] distilled eight commonly used design principles that they found useful in AR. These design principles concern affordance, cognitive overhead, physical effort, learnability, user satisfaction, flexibility in use, responsiveness, feedback, and error tolerance. More recently, Endsley et al. [13] generated eight heuristics with an iterative process working closely with experts and designers. These eight heuristics include fit with user environment and task, form communicates function, minimized distraction and overhead, adaptation to user position and motion, alignment of physical and virtual worlds, fit with user's physical perceptual abilities, and accessibility of off-screen objects. Besides these general-purpose guidelines, researchers have also proposed guidelines for specific platforms (e.g., mobile computing [32]), applications (e.g., education [29]), and user groups (e.g., people with low vision [69, 70]).

2.2 Free-hand AR/VR Interactions

Immersive user interactions are a key aspect of AR and VR. To support interactions between users and digital content, many AR/VR devices rely on controllers. However, controllers not only are cumbersome to carry, but also break immersion, which ultimately makes interactions feel less natural and fluent [38, 39]. To circumvent this issue, researchers have been looking into controller-free interaction methods that leverage the expressivity of users' hands. There are a wide variety of input modalities to enable controller-free interactions, such as gaze [48] and voice [20]. In this section, we focus on ones that leverage users' hands (i.e., free-hand interactions).

Much effort has been spent on enabling free-hand interactions in AR/VR scenarios. In the consumer domain, there are products (e.g., Leap Motion [61], HoloLens [43], and Oculus Quest [40]) that use computer vision to track hands in close range. It is ideal for AR/VR headsets since they often have vantage points that are close to and have clear views of users' hands. In the research domain, people have been exploring alternative and complementary approaches to CV to improve sensing performance using e.g., structured laser beams [31], bio-acoustic vibrations [25], active ultrasonic sensing [44], and skin-mediated radio frequency [68].

With hand tracking, prior research has designed interaction techniques around the expressivity of users' hands. For example, *Open Palm Menu* [2] proposed a series of menus that follow the user's palm of the non-dominant hand, the state of which controls the state of the menu (e.g., the user controls the rendering of the menu by opening or closing the hand). *Plane, Ray, and Point* [26] allows users to create shape constraints by using symbolic gestures to enable precise spatial manipulations of virtual objects. *Portal-ble* [50] proposed sensing and interaction techniques for users to grasp and manipulate virtual objects in smartphone-based augmented reality. Surale et al. [58] explored seven bare hand interaction techniques for mode-switching tasks in VR. Similarly, Masurovsky et al. [38] investigated the performance of controller-free hand interactions in grab-and-place tasks. Both research yielded quantitative and qualitative results that grounded the benefits of controller-free hand interactions in AR/VR. Additionally, there is a major focus of research in AR/VR text entry. For example, researchers have investigated the performance of users' typing on virtual keyboards [11]. In another example, *PinchType* [15] enables users to type with thumb to fingertip pinches. *VirtualGrasp* [65], which is closely related to Hand Interfaces, allows users to easily and naturally retrieve virtual objects in immersive environments by performing hand poses that people commonly use to grasp these objects in reality. The ingenious leveraging of users' real-world experience has led to rich, self-revealing interaction designs that have a high level of consensus across users.

2.3 Leveraging the User Body as AR/VR Interaction Medium

Closest to our research is prior work that has leveraged the user body as AR/VR interaction media which user interfaces can refer to or reside on. First, the user body (e.g., hands and arms) can serve as spatial references to graphical menus [34] and user interactions [54] to facilitate natural and precise input. Specifically, Lediaeva et al. [34] investigated methods to render graphical interfaces around users' hands and arms for AR/VR inputs. *WatchSense* [54] uses fingers touching on the arm as reference points that open up a rich set of finger gestures. Yan et al. [64] investigated acquisitions of targets rendered physically around users utilizing their sense of space and proprioception. Additionally, prior work has demonstrated leveraging the human body as visual reference to facilitate the recollection of interactions [4, 56].

Researchers have also used body surfaces to host conventional GUIs in 2D. For example, *SixthSense* [37], *Skinput* [25], and *OmniTouch* [24] implemented projection and detection systems to render user interfaces on users' hands and arms. Prior work also investigated the efficiency and usability of skin-mediated user interfaces (e.g., *Its a wrap* [5]). Research has also been conducted on on-body interactions that are not specifically designed for AR/VR applications, but could be easily adapted. For a comprehensive review of literature, we recommend this survey [3]. In all prior systems, users need to straighten their palms and fingers to make planar surfaces, best for hosting displays and touch inputs for GUIs. In contrast, Hand Interfaces do not suppress the 3D expressivity of hands, but instead leverage it in rich AR/VR interactions.

We drew much inspiration from prior research showing several seminal ideas of using the 3D characteristics of the user body for interactions. *Let your fingers do the walking* [66] uses the metaphor of the user's fingers becoming their legs to enable efficient travels in VR. Tsuji et al. [59] proposed a method that allows users to animate 3D models with finger play and hand shadow. In a more general-purpose UI design, *DigiGlo* [8] explored using hands as an input and display mechanism through digital gloves, unifying display and interaction in the context of gaming. Compared with prior systems that have digital content floating in the air, or rendered only on the planar parts of the human body, *DigiGlo* coats the user's hands with interfaces, enabling intuitive control, embodiment, and avoiding split attention. In addition, prior research has shown that the user body can be turned into actuation mechanisms, which we believe is a very important aspect of 3D interfaces. For example, Lopes et al. demonstrated adding Electrical Muscle Stimulation (EMS) to heavy objects (e.g., walls) in virtual reality [35], and using force feedback in a wide variety of mixed reality scenarios to enhance user experience [36]. In both projects, the force feedback provided by EMS can be rendered in 3D to accommodate the forms and mechanisms of many 3D objects in AR/VR.

Finally, users' bodies could be used to provide "self-haptic" or "self-contact", which is the haptic feedback of touching one's own body part [6], and has been proven critical in users' sense of embodiment [7]. Using this technique, researchers have created haptic feedback similar to what users would get from touching screens of 2D user interfaces [33] as well as manipulations of everyday 3D objects [14]. The haptic aspect of these techniques overlaps with ours, but users of Hand Interfaces are aware of the embodiment (i.e., their hands transform into objects) rather than being visual illusioned with retargeting in previous work, opening up new research opportunities in embodiment and uniquely enabling AR applications. In addition, Hand Interfaces leverage the dexterity of users' hands for a wide variety of 3D inputs beyond acting as planar surfaces to provide haptic feedback to touch.

As discussed above, there has been much effort in leveraging the user body, especially hands, to improve AR/VR interactions. However, most prior work has focused on controlling off-body interfaces, using the body as a pointing device (e.g., a new type of mouse), or on on-body interfaces that wrap around the user body, like digital skins or gloves. In comparison, Hand Interfaces transform a user's hands into virtual objects (i.e., 3D interfaces) to

Table 1: Design table that highlights the novelty of Hand Interfaces among prior literature in terms of interface location and modality.

		Interface Modality	
		0D/1D/2D/2.5D	3D
Interface Location	Off-body	GUIs decoupled from a user's body (e.g., pointing) [2, 11, 21, 34, 41, 58, 64]	3D virtual objects that do not share voxels with a user's body (e.g., virtual grasp, retargeting) [14, 26, 35, 36, 38, 49, 50, 62, 65]
	On-body	GUIs that reside on a user's body (e.g., digital glove) [5, 8, 24, 25, 37, 68]	3D virtual objects intentionally embodied by users **Hand Interfaces**

Figure 5: Designs that were conceived for the *Spray can* during brainstorming. We considered shape similarity, kinematic similarity, comfort, and social acceptance to pick out the best design, which in this case is the fourth gesture.

interact with the virtual and the physical world in both VR and AR. Moreover, the transformation is transparent to users, uniquely providing users an experience of embodiment [30, 52]. In this paper, we will demonstrate how our interaction techniques may benefit future AR/VR in a wide range of scenarios. Table 1 summarizes the novelty of Hand Interfaces among prior literature on free-hand interactions for AR and VR.

3 INTERACTION DESIGN

Overall, we believe in several innate advantages of Hand Interfaces, which we leveraged when designing our interaction techniques. First, similar to other free-hand interactions, Hand Interfaces are readily available to users, and thus are low-friction in task switching. This is a useful advantage, especially in AR scenarios where users often need their hands to perform tasks in the physical world. Additionally, Hand Interfaces offer tactile feedback by nature due to skin contacts. Finally, Hand Interfaces leverage proprioception, which improves the precision of 3D manipulations. Compared to conventional free-hand interactions, Hand Interfaces can be more expressive in some cases. For example, techniques that leverage grasping gestures to allow users to directly "grasp" virtual objects yield ambiguities in retrieving objects when different objects might have similar grasping postures. This is a typical occurrence since universal industry designs improve the affordance and usability of everyday objects [47]. For example, users grasp a virtual stapler in a way similar to how they might grasp a wrench, a fishing rod, a billiard cue, or any other objects that features a pole-like user interface. In our design process, we aimed to circumvent this issue for Hand Interfaces. To achieve this, Hand Interfaces, at times, compromised other design considerations such as range of motion and flexibility, comfort, or realism. Therefore, it requires a careful design to balance a wide range of design considerations. We discuss this design process in this section.

We conducted concept-driven brainstorming with all researchers contributing to this project. Researchers were asked to come up with designs that use hands to imitate objects as user interfaces. Early ideas mostly involved digital interfaces inspired by prior work and conventional computer platforms. Such designs include using palms as keyboards, making an "O" gesture with the thumb and the index finger as a click wheel, or using the index finger as a slider. We then transitioned to considering 3D objects. We started with common controllers that can be found in the real world, such as a joystick imitated by a thumbs-up gesture (with the thumb "becoming" the stick) and a toggle switch imitated by the index finger joint. Finally, we thought of objects that were less of an interface by design but more-so props that users could utilize in AR/VR environments. Examples include tools such as a mug, hourglass, spray can, a pair of scissors, wrench, fork, ladle, lever, binoculars, and fishing rod; musical instruments such as a trumpet, bongo, and kalimba; educational props such as a globe, magnet, multi-meter probes; and entertainment props such as a wand, color palette, and water gun.

As we designed hand postures that resemble these objects, we incorporated several design considerations, which we list below. We use our design process for *Spray can* (Figure 5) as an illustrative example.

(1) First, we considered **shape similarity**. We removed design ideas with hand poses that least resembled the target objects in shape. As we later found in the study, considering shape similarity contributed to users' perception of realism.
(2) Another consideration we adopted was **kinematic similarity**. In this category, we estimated how similar the dynamic characteristics (e.g., degree of freedom) of hands and objects were. In other words, we expected the motion of rigged virtual objects to properly map to that of a user's hand (or at least a part of it). For instance, in the *Spray can* example, we removed designs that prevent "pressing-in" motions because a spray nozzle is to be pressed. In practice, we found that considering kinematic similarity was critical to user comfort.
(3) We also considered **comfort**. Specifically, we avoided hand poses that were difficult to perform or uncomfortable to maintain. We also removed designs in which the imitating hand is being pressed, bent, or pulled by the manipulating hand in an unnatural way. Though we found consistency in terms of comfort among researchers in this project, we are aware that comfort is highly subjective to individual differences, which should not be overlooked and merit further investigations.
(4) Finally, we considered **social acceptance**. Hand gestures should be socially acceptable as AR/VR has a wide range of applications involving users with different social and cultural backgrounds sharing the same physical spaces. In the *Spray can* example, we removed the design with a curled middle finger for the consideration of social acceptance. Similar to the design consideration of comfort, researchers reached a consensus on social and cultural acceptance of gestures

used in this paper. However, cultural differences should also always be considered outside the paper. Gesture sets should be adapted for target user groups when Hand Interfaces are applied.

Eventually, our design process yielded 28 Hand Interfaces shown in Figure 1. On a high level, these designs can be categorized by the number of hands involved. 22 designs involved one hand and the rest involved two hands to imitate target objects. Specifically, 10 out of 22 one-hand designs required manipulations from the other hand. Interaction techniques of these imitated objects were self-revealing in most designs. For example, to use the *Ladle* and the *Fork*, users should manipulate the objects as they would in reality. Similarly, users raise the *Binoculars* up close to their eyes to transition to a long-range view. A tad more complicated were single-hand designs that involved the other hand for manipulations, which are described below:

- *Joystick* interface (Figure 1*a*) required users to grasp the thumb of the imitating hand (the stick of the joystick) and move it around to control things, e.g. the orientation of a game character.
- *Fishing rod* (Figure 1*e*) was imitated by pointing the thumb horizontally to one side as the reel and extending the other fingers as the rod body. By rotating the thumb with another hand, users were able to wind up a fishing line to reel in their bait.
- *Lever* (Figure 1*i*) was imitated by extending the index and middle fingers. By pinching and moving the two fingers, users could manipulate the end of the lever.
- *Kalimba* (Figure 1*j*), also known as "thumb piano", turned the four fingers (index, middle, ring, pinky) of the imitating hand into four piano keys. Users could tap their fingers to tap the virtual piano keys of the kalimba and create a simple melody.
- *Inflator* (Figure 1*n*), also known as "manual air-pump", was imitated by a spider-man hand gesture. By squeezing the index and the pinky fingers towards each other, users were able to compress the inflator and use it to inflate virtual balloons.
- *Globe* (Figure 1*o*), a universally spherical object, was imitated by a fist. Users could interact with the globe by touching the fist with the index finger of the manipulating hand. Once users clicked on the globe, an enlarged map of the touched location would be displayed.
- *Trumpet* (Figure 1*y*) was imitated by a fist with an extended pinky finger. The pinky finger of the imitating hand represented the bell-like shape of the trumpet and the other fingers represented the trumpet body, with each joint imitating a valve.
- *Toggle switch* (Figure 1*z*) is a switch rendered on the first joint of the index finger of the imitating hand. A user could click on the joint to toggle the lights on and off.
- *Spray can* (Figure 1α), was represented by a hook-like hand gesture where the index finger acted like the nozzle and the other fingers imitated the spray can body. By pressing and holding the index finger of the imitating hand, users were able to spray paint in the air and create 3D artwork.
- *Stapler* interface (Figure 1β) consisted of a base imitated by extending the thumb and a handle imitated by extending the other fingers. Users could bind virtual documents with it by pushing down the handle to the base with the other hand.

For demonstrations of these Hand Interfaces in action, please refer to the Video Figure. Due to the diversity of our gesture set, Hand Interfaces can be used for retrieving objects in AR/VR. Specifically, users can perform a hand pose to retrieve the corresponding virtual object for further interactivity, similar to the retrieval technique demonstrated in *VirtualGrasp* [65]. We have implemented a detection pipeline using Oculus Quest to demonstrate the feasibility of Hand Interfaces. Next, we will discuss our implementation process.

4 IMPLEMENTATION

Hand Interfaces were built on commercially available hardware with custom-designed detection algorithms based on existing hand tracking APIs. In this section, we describe the core components.

4.1 Hardware

We implemented Hand Interfaces with an Oculus Quest (first generation), which was connected to a PC (with an AMD Ryzen 7 3700 CPU and an RTX 3070 8G GPU) using a USB 3.0 Type-C cable. Four built-in cameras on the Quest enabled hand tracking in real-time. We used two Oculus Touch Controllers to implement baseline designs in our user studies.

4.2 Software

The Hand Interfaces software was built using the Unity game engine (2020.3.7f1 version). The detection pipeline is summarized in Figure 6. First, the system initializes at the state of "No hands" before Oculus hand tracking finds any hands in the field of view. Once hands are found, our software transitions to the "Free hand" state. In this state, Oculus hand tracking returns the positions of all hand key points at $\sim$ 35 FPS (measured under our hardware setup and default Oculus+Unity settings). With this data stream, our back-end algorithms continuously compute the similarity between the current hand pose and all gesture templates in the current application's gesture set. The i here refers to any gesture template in the dataset, which can either be designed by authors of the application in advance or defined by users during run time. The N in Figure 6 denotes the number of gesture templates in the dataset. The difference score S_i is calculated by summing the Euclidean distances between each one of the 25 key points on the current hand gesture with its corresponding key point on the i^{th} gesture. Specifically,

$$S_i = \sum_{j=1}^{25} \sqrt{(d_{ij} - c_j)^2}\ , \quad \text{for } i = 1, 2, \ldots, N \tag{1}$$

Note that the more similar the current gesture is to a gesture template, the smaller the difference score is.

Our system tracks the minimum difference score $\min_{i=1,2,\ldots,N} S_i$ in the gesture map. If $\min S$ is higher than an empirically tuned threshold, the software stays at the "Free hand" state. A small threshold requires users to perform a gesture more accurately, while a large threshold may lead to a false positive. Once $\min S$ falls below

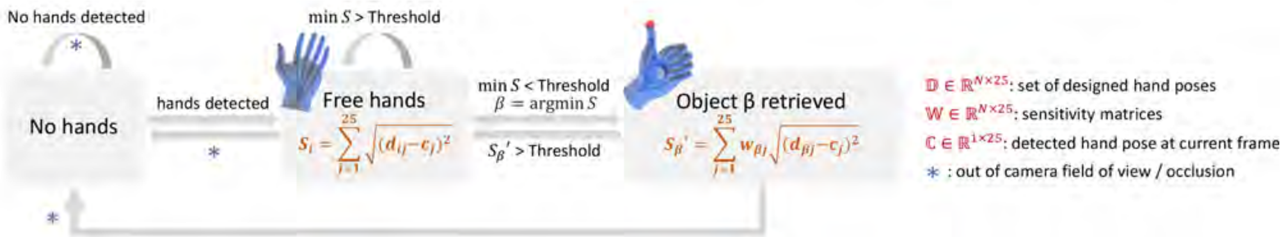

Figure 6: The detection pipeline of Hand Interfaces. Once the hands are found, the algorithm detects whether a user is doing a registered gesture using template matching. When a registered gesture is detected, corresponding virtual objects will be retrieved.

the threshold, the software transitions to the "Object retrieved" state as the object/gesture β is retrieved/detected where $\beta = \text{argmin } S$.

When the current state is "Object retrieved", users can interact with the current virtual object. Along with the thresholding mechanism, we introduce a weighted distance score S'_β to accommodate movements of fingers and joints during users' manipulations of imitated virtual objects with rigged parts. This weighted distance score prevents our gesture detection from falsely recognizing a manipulated hand as a different hand gesture. Note that S'_β is a modified version of S_β. To get S'_β, we simply add weights to S_β's calculation using a sensitivity matrix W_β. Specifically,

$$S'_\beta = \sum_{j=1}^{25} w_{\beta j} \sqrt{(d_{\beta j} - c_j)^2}\ , \quad \beta = \underset{i=1,2,\ldots,N}{\text{argmin}} \ S_i \tag{2}$$

Note that $w_{\beta j}$ is the j^{th} element of W_β, assigning a weight value to the j^{th} key point of current hand pose. W_β is one of the N sensitivity matrices of W. Each sensitivity matrix is tuned for a specific design of Hand Interfaces.

Figure 7 shows an example of tuning a sensitivity matrix for the *Joystick* hand pose skeleton. Specifically, the sensitivity matrix assigns lower weights for rigging key points that are more likely to move (w_4, w_5, w_{19} in the Figure 7), and higher weights for relatively static key points (all the other points in the Figure 7). In pilot experiments, we found that using weighted S'_β is more robust than using the unweighted S_β during the manipulations of imitated objects. When S'_β is greater than the threshold, the software dismisses the object (*Joystick* in this example) and transitions back to the "Free hand" state.

Once objects are retrieved, our software tracks multiple key points on the imitating hand for positioning and orienting the virtual objects. Specifically, the bottom of the palm determines where the object is. Other key points decide the orientations of virtual objects and, in some designs, positions of their parts (e.g., the intermediate and proximal phalange bones control where the keys are in the *Kalimba* design). For Hand Interfaces that involve both hands to imitate objects (e.g., *Book*) we duplicate the hand tracking and heuristics for the other hand. For Hand Interfaces that require the other hand to manipulate virtual objects, we detect touch by tracking Euclidean distances between key points of the manipulating fingers and the imitated object parts that are supposed to be moved around. Below we describe our detection methods grouped by interaction designs:

- For designs that rely on discrete hand poses (e.g., *Scissors*), we use the same software described above. For example, once the angle between the index and middle fingers is smaller than a threshold in the *Scissors* hand pose, the action of cutting is performed.
- For designs that involve proximity-based interactions, we threshold the distances between the anchoring points on the imitating hands and those on objects in the environment. For example, the *Binoculars* design presents users with a long-range view when they raise their hands up close to users' eyes.
- For designs that users click (e.g., *Globe*, *Toggle switch*), we continuously monitor the distance between key points of the manipulating hand and key points of the imitating hand. We detect clicks by looking for patterns of "approach, touch, and release". We set the distance threshold for distinguishing between touch and no touch to 7 mm.

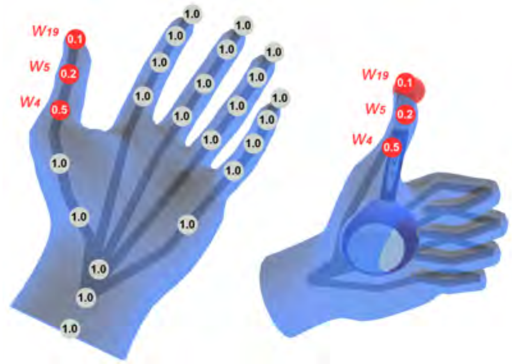

Figure 7: The mapping between the sensitivity matrix for the *Joystick* and its hand pose skeleton in Hand Interfaces. Grey dots indicate static points with default weights of 1. Red dots indicate moving points (thumb proximal phalange bone w_4, thumb distal phalange bone w_5 and the tip of the thumb w_{19}) with reduced weights. Sensitivity matrices enable robust detection during object manipulation.

- For designs that feature handle-like interactions (e.g., *Joystick* and *Fishing rod*), we monitor clicks as well as patterns of "approach, grasp, and move". Designs that rely on complicated manipulations of hands such as *Inflator* and *Spray can* involve heuristics consisting of simpler ones above.

We fine-tuned thresholds for the difference score and sensitivity matrices of each design for robust and precise detection. We implemented a hysteresis buffer of 200 milliseconds of retrieval results to remove jitters to further improve robustness. Furthermore, we cached the latest frames of Oculus hand tracking, which we used to replace current frames in order to prevent transient freezing in hand tracking during brief occlusions.

4.3 Open Source

To facilitate people using Hand Interfaces, we commit to releasing all the design files and source code in an open-source repository: https://github.com/handinterfaces/Hand-Interfaces. Additionally, we will post the research overview, design details, core implementation, Q&A, and demos to our project website: https://handinterfaces.com.

5 STUDY OF OBJECT RETRIEVAL

As prior literature suggested, object retrieval is an important building block for AR/VR interactions [21, 49, 57, 63, 64, 67]. Hence, we investigated users' observation of Hand Interfaces in object retrieval scenarios. We selected 11 designs to use in the study to ensure a completion time of around half an hour. The chosen 11 designs covered all categories across the Hand Interfaces design space, including two-hand designs (*Binoculars*), one-hand designs without manipulation from the other hand (*Scissors*, *Wand*), and one-hand designs requiring clicking/grasping/squeezing from the manipulating hand (*Joystick*, *Fishing Rod*, *Kalimba*, *Inflator*, *Globe*, *Trumpet*, *Toggle Switch*, *Spray Can*). These objects can be found in Figure 8 left. Figure 1 shows hand poses that retrieve these objects.

5.1 Baseline Designs

To best tease out users' observations on Hand Interfaces, we designed two baseline conditions in the retrieval tasks. The first was a drop-down menu with miniaturized objects on a flat 2D plane in a virtual 3D environment, which closely resembles GUIs users are familiar with (Figure 8 left). The models rotated slowly around themselves on the menu for improved visibility. Users used a pair of controllers, with one controller acting as a pointer device. The button underneath the index finger was used to confirm a selection. The other controller acted as an anchor point where virtual objects would be rendered once selected. For the rest of the paper, we refer to this interaction technique as DM (i.e., Drop-down Menu).

The second baseline condition was inspired by *VirtualGrasp* [65] (Figure 8 middle), which allowed participants to use their hands to grasp virtual objects. We designed these baseline gestures based on *VirtualGrasp* and designs that received consensus among researchers. Once certain hand poses were performed, the corresponding virtual objects were rendered on the hands, as if the participants were grasping the virtual objects. For the rest of the paper, we refer to this interaction technique as VG (i.e., *VirtualGrasp*), and Hand Interfaces as HI. An identical set of visual designs were used across interaction techniques for consistency. For implementing DM, we detected where the controller pointing ray intersects the menu using simple ray cast detection. We used the same detection pipeline (Figure 6) for VG and HI with different sets of gestures.

Figure 8: In this study, we explored three techniques for the object retrieval scenarios, including two baselines and hand interfaces. From left to right: Drop-down Menu, VirtualGrasp, and Hand Interfaces.

5.2 Evaluation Configurations

The study was conducted in a quiet lab environment moderated by two experimenters. Participants were seated comfortably in a large chair throughout the study and were free to rest at any point of the study. Audio and video recordings were captured via a GoPro camera to enable reviewing comments made by participants. During the study, the user's headset display was mirrored onto a computer monitor to ensure that the procedure was being followed correctly.

5.2.1 Participants. The user study consisted of 17 participants (9 Females) with ages ranging from 19 to 39 (Mean = 24.1 SD = 4.4). We collected age, gender, education level, major, VR experience level, handedness, and hand size information from participants before the study started. Overall, 9 participants had experience with VR headsets, typically using an Oculus Quest or HTC Vive. 16 users were right-handed and 1 user was left-handed. The average hand width (i.e., from the outer side of the thumb to the outer side of the pinky finger) and length (i.e., from the tip of the middle finger to the base of the palm) measured 15.6 and 18.9 cm, respectively.

5.2.2 Procedure. First, we introduced each of the retrieval techniques and a brief tutorial on how to use the Oculus Quest headset and controllers. Each participant would then be loaded into one of three virtual environments, each corresponding to a different retrieval technique. In each virtual environment, users were tasked to retrieve all 11 virtual objects through either gestures or a menu, depending on the retrieval technique. After retrieving each object, users were asked to give scores on a 7-point Likert scale regarding five metrics, which will be discussed in the following section. The collected data was later used for quantitative analysis. Additionally, participants were asked to provide their rationales and any feedback they might have on each metric and design. The resulting responses were later used for qualitative analysis. Upon answering every question, users proceeded to the next object.

After retrieving all objects in a virtual environment, participants were loaded into the next environment to repeat the same process until they finished every task in all three environments. The order of retrieval techniques and object retrieval were counterbalanced. On average, participants completed the study in half an hour. As per COVID-19 regulations, masks were worn at all times throughout the study and participants were provided hand sanitizer. The headset and controllers were cleaned after each user study session.

5.3 Evaluation Results

We set off to find patterns and consensus among participants' perceptions of freehand interaction techniques for retrieval. To achieve this, we analyzed the quantitative data using clustered boxplot visualization and significance analysis methods, as well as ran a thematic analysis with quotes from 17 participants.

5.3.1 Quantitative Feedback. Participants were asked to provide scores on a 7-point Likert scale on how well the design performed according to certain metrics, with 1 being "strongly disagree" and 7 being "strongly agree". In this study, we used five metrics including, fidelity of retrieval, freedom of movement during retrieval, swiftness of retrieval, comfort of retrieval, and ease of recollection. Fidelity of retrieval refers to the degree of which the interaction technique provided feedback such that the virtual experience was similar to retrieving the object in reality. Freedom of movement refers to how unrestricted participants' movement was during retrieval. Swiftness of retrieval metric quantifies how quickly participants could retrieve an object from the standby hand pose (i.e., relaxed hand). Comfort of retrieval refers to how physically and mentally comfortable participants felt during retrieval. Finally, ease of recollection describes how easily participants could recall how to retrieve objects.

55 data points were collected per user (i.e., 5 data points × 11 interfaces) resulting in 935 data points in total across the 17 users. Figure 9 is a clustered box plot showing distribution, where each box represents 17 samples per metric per retrieval technique. For example, the first box on the left in Figure 9 shows the distribution of 17 samples on fidelity in the Drop-down Menu retrieval technique, with each sample denoting an averaged score across 11 interfaces of a participant.

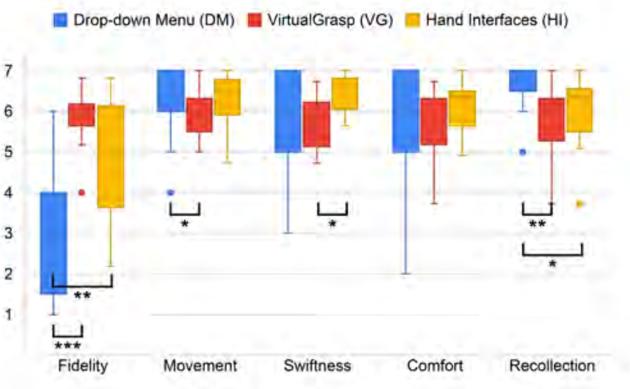

Figure 9: Retrieval Evaluation of Drop-down Menu, VirtualGrasp, and Hand Interfaces. The first (blue), second (red) and third (yellow) box plots in each cluster indicate score distribution of DM, VG, and HI techniques, with 1 being "strongly disagree" and 7 being "strongly agree" on a 7-point Likert scale. The number of asterisks denotes the degree of pairwise significance.

Using the Likert scale data, we conducted a significance analysis among the three retrieval techniques (DM, VG and HI). Due to the non-normal distribution, the ordinal nature of Likert Scale data, the number of categories (which is three instead of two), and the independence among categories, the Independent-Samples Kruskal-Wallis Test, a rank-based nonparametric test, was used to determine differences of statistical significance. The significance level in our test was set to 0.05. When significance was detected (i.e., $p < 0.05$) among three techniques, we further conducted pair-wise comparisons using asymptotic significance values adjusted by the Bonferroni correction (i.e., p_{adj}) for multiple 2-sided tests. Figure 9 visualizes different degrees of significance using asterisks ("$*$" denotes $p_{adj} < 0.05$, "$**$" $p_{adj} < 0.01$, and "$***$" $p_{adj} < 0.001$). In retrieval tasks, with regard to fidelity, we found HI and VG to be significantly better than DM with $p_{adj} < 0.01$ and $p_{adj} < 0.001$, respectively. We did not find a significant difference between HI and VG. The movement metric results indicated that HI had no significant differences against DM and VG. The swiftness metric showed that participants preferred HI to VG with a significant difference of $p_{adj} < 0.05$. There were no significant differences in comfort among the three interaction techniques. Finally, results indicated that it was easier to recall how to retrieve objects with DM than with VG and HI with $p_{adj} < 0.01$ and $p_{adj} < 0.05$ respectively, while we found no significant difference between VG and HI.

In addition to the overall comparison between HI and the two baseline techniques, we probed deeper into each virtual object design, as we acknowledge that participants' perceptions of interaction techniques might vary from design to design. We examined design-wise population percentages for all metrics and all interaction techniques, which can be found in the appendix. Here we use the fidelity metric as an example. Figure 10 shows retrieval fidelity scores for VG and HI with a list of object designs (i.e., interfaces) as the vertical axis and population percentage as the horizontal axis. Seven distinct colors represent scores from 1 to 7 in the Likert scale. The length of the colored bars reflects participant percentages.

This visualization reveals the differences across designs – not only in score averages, but also in divergences. Additionally, participants' feedback on HI was more divergent than VG in general. This divergence was also reflected in participants' qualitative feedback, which is discussed below. In HI, the three designs that received the most positive feedback – *Globe*, *Scissors*, *Binoculars* involved gestures people commonly make in real life, which helped with participants' perception of realism. This observation was confirmed by our qualitative result analysis.

5.3.2 Qualitative Feedback — Real-world Experience on the Perception of Realism. People's real-world experience affects their perception of realism in AR/VR, which we found to be user-dependent. In this section, we report the various kinds of real-world experience that participants reported.

For VG, participants' perception of realism when grasping certain objects depended on 1) whether participants have grasped those objects before in reality, 2) whether participants knew how to grasp those objects, and 3) whether participants were able to grasp those objects the same way in our study compared to in real life. For objects that people normally would not grasp (e.g., the *Toggle switch*), participants yielded lower scores. Additionally, participants who did not know how to hold a Kalimba in reality reported that they could not determine whether the *Kalimba* was realistic. This lack of real-world experience also contributed to reducing the ease of recollection, as we found that 4 out of 17 participants commented that their perception of realism affected the ease of recollection. Furthermore, we saw divergence in participants' feedback. When the hand poses matched how people would grasp objects in reality, participants tended to perceive them as more realistic than if the poses did not match. Two example quotes include P9 "This is not how I would actually hold a wand", and P1 "It is how we actually hold a trumpet in reality".

For HI, we found that participants applied their experience in performing certain hand poses to the perception of realism. For example, P17 commented on the *Scissors* that "I feel the scissors are realistic because I am familiar with the rock paper scissors poses," which four additional participants agreed with. Other examples similar to this include the *Wand* and the *Binoculars*, both of which feature hand poses commonly used in the real world. Participants did not focus on whether the hand poses matched how they hold objects in reality, as they knew that HI was intended to allow them to imitate the object instead of grasp it. Problems surfaced when participants had a difficult time imagining their hands turning into objects. In fact, "imagine", "think of it" or other similar terms were frequently mentioned in cases where participants found HI realistic. For example, P13 commented: "Sphere looks weird because it is like my hand is wrapped with the globe. However, if I think it of my hand becoming a globe, it seems very realistic. It depends on how I think about it." However, in the case of Kalimba, both P1 and P16 commented that it was realistic because it was easy to imagine their fingers as keys on a Kalimba.

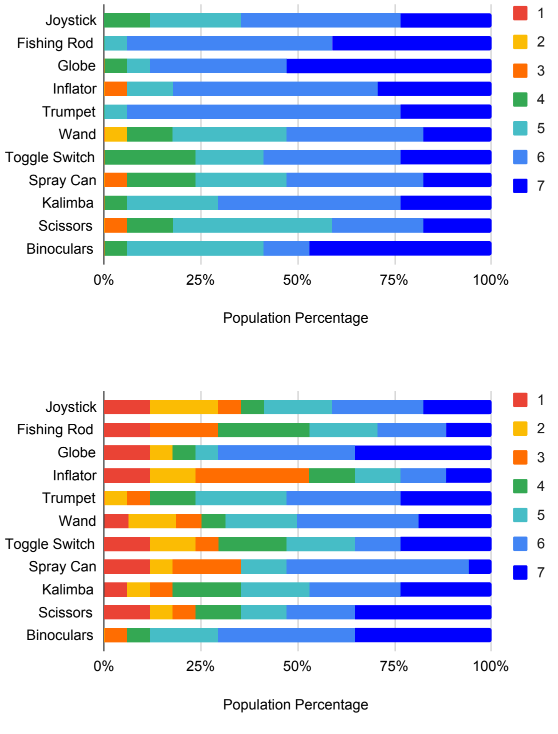

Figure 10: Score percentages of VirtualGrasp (top) and Hand Interfaces (bottom) per Interface on Fidelity. Seven colors denote a score on the mentioned 7-point Likert scale. The length of each color bar illustrates the number of users giving each score. The percentages are shown on the horizontal axis, while interface-wise user preferences are visible across each row.

5.3.3 *Qualitative Feedback — Visuals on the Perception of Realism.*

We found that visual feedback also played a major role on how realistic designs were to participants. In particular, we found that the visual discrepancy between participants' hands and the virtual objects had a major impact on the perception of realism. This effect showed up in both VG and HI. Specifically, in VG, participants expected virtual hands to not clip through objects, while in HI, they expected virtual hands to perfectly align with objects.

For VG, we saw visuals being one of the most commented criteria participants used to assess how realistic designs were. For example, participants paid attention to how well their fingers lined up with the contours of objects. Seeing virtual hands clipped through objects due to detection and rendering imperfections resulted in a considerable negative opinion about the interaction. Interestingly, we observed something similar to the uncanny valley effect — we found that discrepancies between users' mental model and what they saw, especially the minute ones, could severely impact the perception of realism. This effect, in some cases, made VG less favorable compared to HI because participants often came with a well-established mental model of how things should be grasped (i.e., VG), but not necessarily imitated (i.e., HI).

For HI, visual discrepancy occurred when there were offsets between the hands and virtual objects they were supposed to imitate. For example, participants disliked the offset between the *Fishing rod* and their index fingers. Similar complaints about offsets were also made by participants in the *Wand* scenario where the wand was slightly longer than their index fingers. Several participants gave low scores for fidelity solely because of seeing hands overlapping with objects, which contradicted their experience with the physical world. One example of this, quoted from P12, said that "Hand interfaces are not realistic because it feels odd for objects to go through my hand". Some participants suggested that we hide the virtual hands to mitigate this conflict for easing first-time users in. In the study, we visualized hands for all interaction techniques for consistency. However, we believe that HI holds a unique advantage by allowing for the virtual hands to be hidden, in which case virtual objects will serve as visual cues for users' hands. For example, during piloting, one participant commented that the *Kalimba* could still be easily used if virtual hands disappeared once the object retrieval is completed.

5.3.4 *Qualitative Feedback — Ease of Recollection.*

As noted previously, we found that the lack of real-world experience negatively influenced the ease of recollection, as 4 out of 17 participants had comments that linked ease of recollection with a perception of realism. In other words, if participants have never used or seen certain objects, these objects were more difficult to remember than others.

For VG, we found ambiguity to be a major factor that affected the ease of recollection. Ambiguity refers to cases where there were multiple different hand poses to grasp an object or the same hand pose could be used to grasp multiple different objects. Six participants commented on this. For example, P12 mentioned that the hand pose to retrieve the *Wand* was too similar to the one used for the *Fishing rod*, making both difficult to recall. Similarly, P16 found that the hand pose to retrieve the *Toggle switch* was similar to how people grasp many small objects, therefore making it difficult to recall. Lastly, P13 mentioned that they might grasp a Kalimba with a different hand pose in real life, and therefore it was difficult to recall.

For HI, we received no comments regarding ambiguity. However, we acknowledge that this might be due to our selection of designs for the study. Nonetheless, we believe that ambiguity is less of a problem when using HI than VG, because there are less variations of grasping gestures than shapes of objects that our hands can imitate. However, having more variations impedes recollection. For example, one participant expressed concerns that because there are too many objects in the real world, we may not be able to remember every interface that are available (P4). Overall, our observation is that having too many gestures to remember can be just as difficult as having very similar gestures to remember. Because each technique has unique pros and cons that compensate for one another, it would be beneficial to consider both in AR/VR interaction designs.

5.3.5 Qualitative Feedback — Comfort. Finally, we discuss comfort in the last section. Four participants found several hand poses in VG uncomfortable. Among these participants, three found the needed twist of specific fingers (i.e., pinky in *Trumpet* and *Spray can*) or wrist (i.e., *Kalimba*) to be awkward to perform. Two participants mentioned that the *Spray can*, *Joystick*, and *Inflator* could not be maintained for a long time due to fatigue. We did not notice significant differences between participants' quantitative feedback on the comfort of any designs. However, comfort levels were more design-dependent in HI than in VG. In other words, future HI techniques need to be designed more carefully to avoid uncomfortable finger/wrist angles that might result in strain and fatigue.

6 STUDY OF INTERACTIVE CONTROL

With the same set of participants and virtual objects, we conducted a second study to investigate users' perception of Hand Interfaces in interactive control scenarios.

6.1 Baseline Designs

We included two baseline techniques in this study. With the first baseline technique, participants performed a fist gesture as an anchor in free space to move virtual objects around for interactions (Figure 11 left). We selected this gesture because we found the fist gesture to be one of the simplest gestures for most people to perform. This gesture simply served as an anchor point, without considering objects' shapes or rigging mechanisms. For the rest of the paper, we refer to this technique as FG (i.e., Fist Gesture).

The second baseline technique was inspired by many educational apps and games in VR (e.g., [23]), as well as prior work (e.g., [49]); users interact with virtual objects by directly manipulating them with their hands (Figure 11 middle). This direct manipulation technique includes moving parts of objects (e.g., *Joystick*), sensing touch (e.g., *Globe*), changing shapes (e.g., *Inflator*), and more fine-grained controls. For the rest of the paper, we refer to this technique as VM (i.e., Virtual Manipulation).

Note that we did not use hand controllers when implementing the baselines to focus participants on aspects that were close to our contributions against prior hand-free systems as opposed to mechanisms that are innately challenging to accomplish using controllers. For example, for the "rotating a joystick" task, the corresponding controller-based method would be clicking the index trigger on the controller of the manipulating hand to rotate the red stick around, which is far from real-world experiences. A baseline with hand controllers would yield results that overlap with studies in prior works on free-hand interactions, and was thus omitted in our study.

6.2 Evaluation Configurations

This study was conducted after the object retrieval study and involved the same set of participants. To avoid technical performance affecting users' perception of interaction techniques, we removed object retrieval in this study. Specifically, the study started with virtual objects already coupled with users' hands. Experimenters switched between the designs by pressing shortcut keys on a physical keyboard.

6.2.1 Procedure. The procedure for interactive control was mostly the same as before in the case of object retrieval. Each participant would be loaded into one of three virtual environments, each corresponding to a different interaction technique. In each virtual environment, users were asked to interact with all 11 virtual objects (some examples can be found in the Interaction Design section). After interacting with each object, users were asked to give scores on a 7-point Likert scale regarding five metrics, the same as the ones in the previous study. This study also took about half an hour to finish.

6.3 Evaluation Results

6.3.1 Quantitative Feedback. Although we used the same set of metrics in this study as the ones used in the object retrieval study, we altered their meanings slightly. Fidelity of interaction refers to

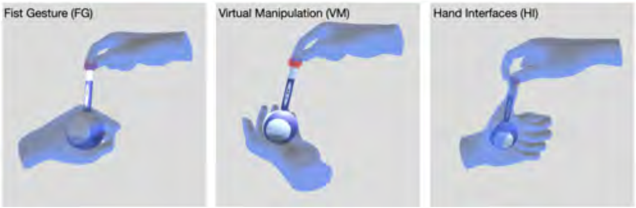

Figure 11: In this study, we explored three techniques for the interactive control scenarios, including two baselines and hand interfaces. From left to right: Fist Gesture, Virtual Manipulation, and Hand Interfaces.

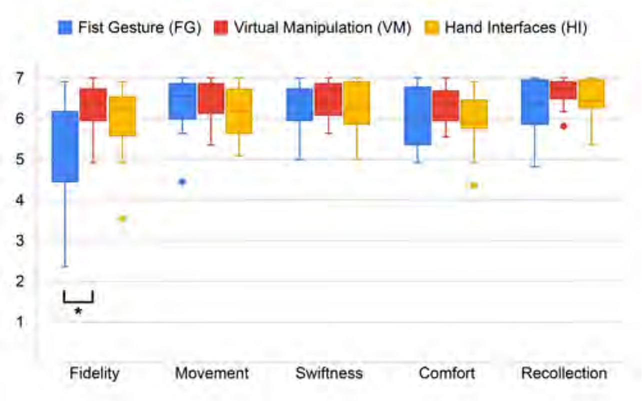

Figure 12: Interaction Evaluation of Fist Gesture, Virtual Manipulation, and Hand Interfaces. The first (blue), second (red) and third (yellow) box plots in each cluster indicate score distribution of DM, VG, and HI techniques, with 1 being "strongly disagree" and 7 being "strongly agree" on a 7-point Likert scale. The number of asterisks denotes the degree of pair-wise significance.

how realistic the interactive control feels to the participants with regard to its dynamic characteristics, which include both the tactile feedback and the visual feedback. Freedom of movement refers to how unrestricted a participant's movement is during the interaction. Swiftness of interaction quantifies how quickly participants can interact with an interface. Comfort of interaction refers to how physically and mentally comfortable participants feel during the interaction. The last metric, ease of recollection, describes how easily participants can recall how to perform interactive control with objects. In total, we collected the same amount of data points as the object retrieval study. Figure 12 shows the result.

As in the previous study, we used the Independent-Samples Kruskal-Wallis Test for significance analysis among the three interactive techniques (FG, VM and HI). The significance level of Kruskal-Wallis Test was 0.05, the same as before. When significance was detected (i.e., $p < 0.05$), we used Bonferroni correction to get adjusted asymptotic significance values (i.e., p_{adj}), which we then used in pair-wise comparisons. We set three pair-wise degrees of significance indicated by asterisks ("$*$" denotes $p_{adj} < 0.05$, "$**$" $p_{adj} < 0.01$, and "$* * *$" $p_{adj} < 0.001$). Figure 12 shows our results from the interactive control tasks.

Referring to Figure 12, VM received a significantly higher average score than FG in interaction fidelity ($p_{adj} < 0.05$). Participants commented that the lack of tactile feedback in FG had a negative impact on their scores. Interestingly, while VM also lacked tactile feedback, the score it received was comparable to HI. This was possibly due to the fact that participants' prior experience of grasping objects in reality contributed to their perception of fidelity. The results indicated that having a realistic grasping gesture contributed similarly to having tactile feedback. We found no significance across the rest of the metrics (i.e., movement, swiftness, comfort, and recollection).

We also examined the design-wise population percentages for all metrics and all interaction techniques, which can be found in the appendix. Figure 13 contains two examples for the movement metric for FG and HI. Figure 13 left, shows that the FG *Fishing rod* received a majority score of 7 by 11 participants while Figure 13 right, shows that the HI *Fishing rod* received a majority score of 5 by 8 participants. This indicated a clear preference for the FG *Fishing rod* over that of HI with respect to freedom of movement. FG interaction, as we expected, was least restricted by the anatomical restraints of the hand among all interaction techniques. In contrast, the HI interaction, which required pinching the thumb and, therefore, constrained the interaction to the degrees of freedom the thumb joint offered, ultimately received a lower movement score. A similar case was exhibited with the *Joystick* scores; however, they were less affected compared to the *Fishing rod* likely due to the fact that the *Joystick* is more kinematically similar to the thumb than a rotating reel.

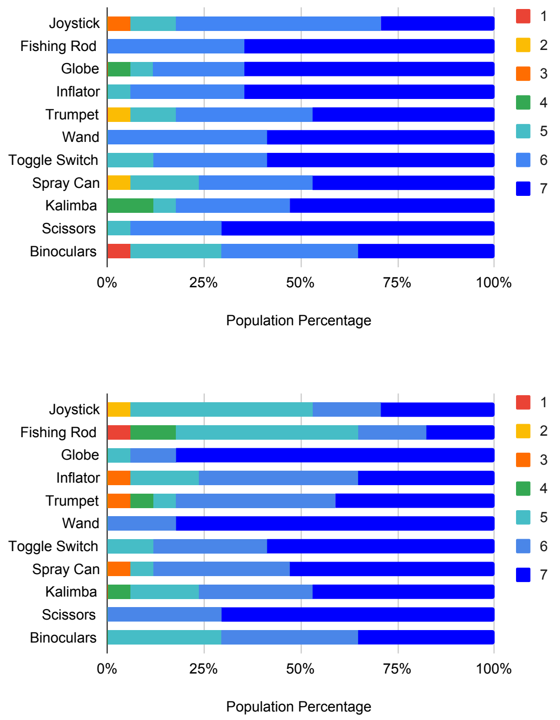

Figure 13: Score percentages of Fist Gesture (top), and Hand Interfaces (bottom) per Interface on Freedom of Movement. Seven colors denote a score on the mentioned 7-point Likert scale. The length of each color bar illustrates the number of users giving each score. The percentages are shown on the horizontal axis, while interface-wise user preferences are visible across each row.

6.3.2 *Qualitative Feedback — Real World Experience on the Perception of Realism.*

Because participants performed interactivity tasks following object retrieval, participants' perception of interactions was inevitably influenced by their prior knowledge gained during object retrieval. Therefore, insights reported in the previous section largely applies to this section. To avoid repetition, we focus on communicating new knowledge generated from interactions (i.e., use of objects) as opposed to object retrieval.

One key finding from the study was around uni-manual vs. bi-manual manipulations. For instance, two participants noted that it felt unnatural to use both hands for objects that people would normally use in a uni-manual way. The *Spray can* is one such example. P16 explicitly commented that the consistency between object manipulation in real world and in VR space had a greater impact on their perception of realism than tactile feedback. From this feedback, we learned that HI designs could be perceived as unrealistic even with the added benefit of tactile feedback if they are not designed in a way that is reasonably consistent with participants' real-world experience.

Overall, we found that tactile feedback has a positive impact on participants' perception of realism. P3 commented when using the *Joystick* that having something tangible (i.e., the thumb) to grasp made it easier for the manipulating hand to maintain its position, which was similar to what they do in the real world. Nine participants, explicitly critiqued VM on the realism front for its lack of tactile feedback. For example, P14 commented that "For switch, there is no tangible feedback, and therefore I think it is not realistic". Similarly, P11 mentioned that "For tasks requiring much interaction, I prefer HI because of its tangible feedback". In our analysis, the average fidelity score of HI in interactive control scenarios was significantly higher ($p < 0.05$ calculated using Independent-Samples Mann-Whitney U Test) than that in object retrieval scenarios, which we suspect was due to the tactile feedback.

6.3.3 Qualitative Feedback — Tactile Feedback.

Participants commented on tactile feedback beyond the scope of perception on realism. Six participants reported various reasons why tactile feedback was beneficial to their interactions with virtual objects. For example, P10 preferred the *Kalimba* design in HI over VM during both imitating and manipulating tasks, as they were able to better detect key presses with the help of tactile feedback. Additionally, P3 felt that controlling the *Joystick* was better done with something tangible so that they could grasp something as opposed to nothing in the air. Furthermore, P11 commented that "the inflator is most comfortable here because when I touch and squeeze my left-hand fingers, it feels easy to keep my hands stable." P11 further stated that "keeping hands stable makes me feel comfortable." Finally, towards the end of the study, we anecdotally found a correlation between tactile feedback and joyfulness of AR/VR designs, as many participants expressed excitement and amusement with a bit of surprise when they first felt the hand-imitated objects in HI by touching them.

6.3.4 Qualitative Feedback — Ease of Recollection.

Once the objects were retrieved, participants naturally found ways to interact with those objects, as they have used many of these objects in the real world. In the study, we found that participants relied solely on appearances to recall the interactions with objects. As a result, the ease of recollection for interaction was more about designing virtual objects with self-revealing or easy-to-remember affordances than anything else regarding the couplings between these objects and users' hands. However, one interesting point we discovered was that participants found interactions with an object to be helpful to the recollections of hand gestures for object retrieval. For instance, P14 mentioned that the *Inflator* hand pose suddenly became comprehensible and, therefore, became easier to recall once they knew how to interact with it.

6.3.5 Qualitative Feedback — Constraints from Hand Kinematics.

Another main discussion point was around the constraints caused by hand kinematics. At the end of the day, our knuckles and joints can only move so much before we experience discomfort. However, the objects we want to imitate may require movements beyond our capabilities. This innate constraint of our hands affects three factors we chose for evaluation — movement, swiftness, and comfort. Two participants noted the discrepancy between limited finger movements and objects in the *Fishing rod* design. One other participant mentioned that this discrepancy slowed down the interaction — the interaction became cumbersome. Participants perceived the *Joystick* as a better design than the *Fishing rod* for the better kinematic similarity between the thumb joint and the expected movements of a joystick. However, two participants suggested that the limited range of motion of their thumbs should be extended with some visual compensation. This insight leads to an intriguing design opportunity of using visual illusion or scaling (e.g., [1, 16, 19]) to compensate for the limited range of motion in HI, and, broadly, our hands. For example, the thumb of a user could be rotated with a much larger angle visually than in reality. Visual illusion and scaling have been proven successful by prior work in compensating for constraints in the physical world (e.g., limited room space) and we expect it to be a promising approach in mitigating some of the challenges we faced in this research.

6.3.6 Qualitative Feedback — Miscellaneous Insights.

Proprioception. When imitating objects with hands, we found that proprioception came into play across several cases, which participants found favorable. For example, HI allowed participants to use their index finger as a wand, which one participant found particularly useful, as this allowed them to point with greater precision than techniques that asked participants to grasp. Latter techniques required users to rely mostly on visuals when pointing, whereas HI leveraged proprioception of users' fingers.

Social acceptance. One participant commented that the hand pose involving a raised pinky finger has social implications and, therefore, should not be used in public. Another participant noted that it is possible to use the middle finger as a joystick, which we felt might be socially unacceptable in most scenarios. These two examples reminded us that social acceptance might be a challenging problem to tackle, and it would be beneficial to consider both application contexts and user backgrounds when optimizing for social acceptance in HI.

Fun to use. Even though we did not explicitly ask participants, two of them stated that HI is fun to use. Specifically, they found their hands morphing into objects (i.e., *Globe* and *Kalimba*) amusing. This inspired us to strategically amplify the morphing effect (i.e., the transition between users' hands and imitated objects) in future designs.

Ambiguity in detection. As we mentioned earlier, many participants noticed the similarity between hand poses when retrieving objects with VG. Not only did this ambiguity demand more effort when memorizing, but it also created challenges in detection. In fact, to make sure the technical performance did not interfere with the participants' perceptions, we had to partition the designs of VG into batches with objects needing similar retrieval gestures never showing up in the same batch. This effort successfully avoided false detection results without compromising the study flow, since the transitions between batches were handled by an experimenter pressing a key on a keyboard. However, ambiguity still remains a challenging problem to resolve with freehand grasping. Conversely, HI often yields a more diverse set of hand poses that makes detection easier and simpler.

Design space / Customization. Finally, participants praised HI for its creativity and ease of use — some even proposed their own ideas of HI. For example, one participant wanted to customize the design for the *Inflator* by forming a "C" shaped hand poses with four fingers on top and the thumb at the bottom and being able to pinch them. Given the expressiveness of our hands, there are virtually infinite possible designs in HI. Overall, we found the customizability and the design space to be rich in HI.

7 DISCUSSION

First, our experiments suggest that a robust detection system requires fine-tuned sensitivity matrices (i.e. having lower weights for key points that move around during interactions), which were heuristically generated based on researchers' practices and piloting results. However, this approach might not be intuitive to future users with little programming experience. To resolve this issue, we envision an automated calibration process to generate sensitivity matrices, which seems feasible given the known kinematic characteristics of human hands and the imitated objects. A toolkit with user interfaces that allow users to easily calculate and adjust parameters in a calibration process should be helpful.

We suspect that design strategies involving embodiment [30] could further improve users' experience with Hand Interfaces. Both imitation and embodiment seek morphological similarities between hands and objects. While this work has investigated imitation by matching hands to objects, prior work in embodiment matches objects to hand (an opposite direction) by adding morphological similarities to objects. In the context of Hand Interfaces, we could render objects with skin textures or modify object shapes so that they are morphologically similar to human hands. We expect future work to incorporate both directions to further improve the usability of Hand Interfaces.

There are technical challenges within the computer vision approach that powers Hand Interfaces' detection. For example, when two hands interact with each other, the resulting occlusion causes errors in hand tracking. To compensate for this issue, participants were asked to perform gestures that avoid occlusions in our studies, which might have introduced bias in users' perceptions. Due to this technical challenge, we specifically skipped the technical evaluation of objective measures such as detection latency and task completion time. Instead, we focused on subjective measures, which we believe were affected less by sensing imperfections and were more aligned with our intended research contributions.

Additionally, Hand Interfaces could not support virtual objects that are too small or too large in comparison with human hands. Future efforts should look into visual illusion and scale (e.g., [1, 16, 19]) to address this limitation. For example, Hand Interfaces could potentially leverage animations of metamorphosis to let users think of their hands as hands of giants to imitate large objects such as a car.

Overall, we believe Hand Interfaces introduce a new way to interact with the virtual world. Existing interaction techniques focus on leveraging design ideas that people have experience with. For example, the idea behind the Drop-down Menu (i.e., the baseline technique in our study) is the same as the GUIs on conventional computer platforms. It leverages the user experience of using these digital interfaces. Meanwhile, direct Virtual Manipulation (e.g., [23, 65]) opens up a new direction, as it attempts to reproduce the real-life manipulation of physical objects (i.e., physical interfaces) in the virtual world. These conventional techniques leverage users' experience with interfaces (i.e., digital and physical) in 3D virtual worlds to improve learnability but may have missed a large design space unique to AR/VR. Hand Interfaces do not primarily attempt to leverage users' experience but, instead, creates a new interaction modality specific to AR/VR by transforming users' hands into objects through imitation.

8 CONCLUSION

We present Hand Interfaces, an interaction technique that allows users to quickly and easily use a wide spectrum of virtual 3D objects in AR/VR environments by using their hands to imitate objects. We have come up with 28 designs around this interaction technique and conducted two main user studies. The first user study investigated Hand Interfaces in object retrieval tasks, and the second study investigated Hand Interfaces in interactive control tasks. Each study included two baseline techniques, which we drew from prior work and existing applications. We collected quantitative and qualitative feedback from 17 participants and the results indicated that Hand Interfaces are effective, expressive, and fun to use. We demonstrated example applications centering around three domains — entertainment, education, and ubiquitous computing. All these efforts have been open-sourced to facilitate future research.

REFERENCES

[1] Parastoo Abtahi, Mar González-Franco, Eyal Ofek, and Anthony Steed. 2019. I'm a Giant: Walking in Large Virtual Environments at High Speed Gains. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019.* ACM, 522. https://doi.org/10.1145/3290605.3300752

[2] Takumi Azai, Mai Otsuki, Fumihisa Shibata, and Asako Kimura. 2018. Open Palm Menu: A Virtual Menu Placed in Front of the Palm. In *Proceedings of the 9th Augmented Human International Conference.* 1–5.

[3] Joanna Bergström and Kasper Hornbæk. 2019. Human-Computer Interaction on the Skin. *ACM Comput. Surv.* 52, 4 (2019), 77:1–77:14. https://doi.org/10.1145/3332166

[4] Joanna Bergstrom-Lehtovirta, Sebastian Boring, and Kasper Hornbæk. 2017. Placing and Recalling Virtual Items on the Skin. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017.* ACM, 1497–1507. https://doi.org/10.1145/3025453.3026030

[5] Joanna Bergstrom-Lehtovirta, Kasper Hornbæk, and Sebastian Boring. 2018. It's a Wrap: Mapping On-Skin Input to Off-Skin Displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018.* ACM, 564. https://doi.org/10.1145/3173574.3174138

[6] Sidney Bovet and Ronan Boulic. 2017. Ensuring Self-Haptic Consistency for Immersive Amplified Embodiment. (2017).

[7] Sidney Bovet, Henrique Galvan Debarba, Bruno Herbelin, Eray Molla, and Ronan Boulic. 2018. The Critical Role of Self-Contact for Embodiment in Virtual Reality. *IEEE Trans. Vis. Comput. Graph.* 24, 4 (2018), 1428–1436. https://doi.org/10.1109/TVCG.2018.2794658

[8] Julia Chatain, Danielle M. Sisserman, Lea Reichardt, Violaine Fayolle, Manu Kapur, Robert W. Sumner, Fabio Zünd, and Amit H. Bermano. 2020. DigiGlo: Exploring the Palm as an Input and Display Mechanism through Digital Gloves. In *CHI PLAY '20: The Annual Symposium on Computer-Human Interaction in Play, Virtual Event, Canada, November 2-4, 2020.* ACM, 374–385. https://doi.org/10.1145/3410404.3414260

[9] Jiawen Chen, Shahram Izadi, and Andrew W. Fitzgibbon. 2012. KinÊtre: animating the world with the human body. In *The 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12, Cambridge, MA, USA, October 7-10, 2012.* ACM, 435–444. https://doi.org/10.1145/2380116.2380171

[10] Diane Dewez, Ludovic Hoyet, Anatole Lécuyer, and Ferran Argelaguet-Sanz. 2021. Towards "Avatar-Friendly" 3D Manipulation Techniques: Bridging the Gap Between Sense of Embodiment and Interaction in Virtual Reality. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021.* ACM, 264:1–264:14. https://doi.org/10.1145/3411764.3445379

[11] John J. Dudley, Hrvoje Benko, Daniel Wigdor, and Per Ola Kristensson. 2019. Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2019, Beijing, China, October 14-18, 2019.* IEEE, 289–300. https://doi.org/10.1109/ISMAR.2019.00027

[12] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. 2007. Applying HCI principles to AR systems design. (2007).

[13] Tristan C Endsley, Kelly A Sprehn, Ryan M Brill, Kimberly J Ryan, Emily C Vincent, and James M Martin. 2017. Augmented reality design heuristics: Designing for dynamic interactions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 2100–2104.

[14] Cathy Mengying Fang and Chris Harrison. 2021. Retargeted Self-Haptics for Increased Immersion in VR without Instrumentation. In *UIST '21: The 34th Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, October 10-14, 2021.* ACM, 1109–1121. https://doi.org/10.1145/3472749.3474810

[15] Jacqui Fashimpaur, Kenrick Kin, and Matt Longest. 2020. PinchType: Text Entry for Virtual and Augmented Reality Using Comfortable Thumb to Fingertip Pinches. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI 2020, Honolulu, HI, USA, April 25-30, 2020.* ACM, 1–7. https://doi.org/10.1145/3334480.3382888

[16] Tiare M. Feuchtner and Jörg Müller. 2017. Extending the Body for Interaction with Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017.* ACM, 5145–5157. https://doi.org/10.1145/3025453.3025689

[17] Frank Spillers and Ditte Hvas Mortensen. 2020. How to Design Gesture Interactions for Virtual and Augmented Reality. https://www.interaction-design.org/literature/article/how-to-design-gesture-interactions-for-virtual-and-augmented-reality Retrieved January 01, 2022.

[18] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021.* ACM, 662:1–662:14. https://doi.org/10.1145/3411764.3445576

[19] Mar Gonzalez-Franco and Jaron Lanier. 2017. Model of Illusions and Virtual Reality. *Frontiers in Psychology* 8 (2017), 1125. https://doi.org/10.3389/fpsyg.2017.01125

[20] Alon Grinshpoon, Shirin Sadri, Gabrielle J. Loeb, Carmine Elvezio, and Steven K. Feiner. 2018. Hands-Free Interaction for Augmented Reality in Vascular Interventions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018, Tuebingen/Reutlingen, Germany, 18-22 March 2018.* IEEE Computer Society, 751–752. https://doi.org/10.1109/VR.2018.8446259

[21] Tovi Grossman and Ravin Balakrishnan. 2006. The design and evaluation of selection techniques for 3D volumetric displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, Montreux, Switzerland, October 15-18, 2006.* ACM, 3–12. https://doi.org/10.1145/1166253.1166257

[22] Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. 2010. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology.* 3–12.

[23] Hand Physics Lab. 2021. Hand Physics Lab on Oculus Quest. https://www.oculus.com/experiences/quest/3392175350802835/?locale=en_US Retrieved January 01, 2022.

[24] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA, USA, October 16-19, 2011.* ACM, 441–450. https://doi.org/10.1145/2047196.2047255

[25] Chris Harrison, Desney S. Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010.* ACM, 453–462. https://doi.org/10.1145/1753326.1753394

[26] Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. Plane, Ray, and Point: Enabling Precise Spatial Manipulations with Shape Constraints. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, UIST 2019, New Orleans, LA, USA, October 20-23, 2019.* ACM, 1185–1195. https://doi.org/10.1145/3332165.3347916

[27] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2018. Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. In *Artificial Intelligence in Education - 19th International Conference, AIED 2018, London, UK, June 27-30, 2018, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 10947).* Springer, 154–168. https://doi.org/10.1007/978-3-319-93843-1_12

[28] Justinmind. 2019. 18 AR, MR and VR design principles. https://www.justinmind.com/blog/vr-design/ Retrieved January 01, 2022.

[29] Hannes Kaufmann and Andreas Dünser. 2007. Summary of Usability Evaluations of an Educational Augmented Reality Application. In *Virtual Reality, Second International Conference, ICVR 2007, Held as part of HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings (Lecture Notes in Computer Science, Vol. 4563).* Springer, 660–669. https://doi.org/10.1007/978-3-540-73335-5_71

[30] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence Teleoperators Virtual Environ.* 21, 4 (2012), 373–387. http://www.mitpressjournals.org/doi/abs/10.1162/PRES_a_00124

[31] David Kim, Otmar Hilliges, Shahram Izadi, Alex Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *The 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12, Cambridge, MA, USA, October 7-10, 2012.* ACM, 167–176. https://doi.org/10.1145/2380116.2380139

[32] Sang Min Ko, Wonsuk Chang, and Yong Gu Ji. 2013. Usability Principles for Augmented Reality Applications in a Smartphone Environment. *Int. J. Hum. Comput. Interact.* 29, 8 (2013), 501–515. https://doi.org/10.1080/10447318.2012.722466

[33] Luv Kohli and Mary C. Whitton. 2005. The haptic hand: providing user interface feedback with the non-dominant hand in virtual environments. In *Proceedings of the Graphics Interface 2005 Conference, May 9-11, 2005, Victoria, British Columbia, Canada.* Canadian Human-Computer Communications Society, 1–8. https://dl.acm.org/citation.cfm?id=1089510

[34] Irina Lediaeva and Joseph J. LaViola. 2020. Evaluation of Body-Referenced Graphical Menus in Virtual Environments. In *Proceedings of the 45th Graphics Interface Conference 2020, Toronto, ON, Canada, May 28-29, 2020.* Canadian Human-Computer Communications Society, 308–316. https://doi.org/10.20380/GI2020.31

[35] Pedro Lopes, Sijing You, Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. 2017. Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017.* ACM, 1471–1482. https://doi.org/10.1145/3025453.3025600

[36] Pedro Lopes, Sijing You, Alexandra Ion, and Patrick Baudisch. 2018. Adding Force Feedback to Mixed Reality Experiences and Games using Electrical Muscle Stimulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018.* ACM, 446. https://doi.org/10.1145/3173574.3174020

[37] Pattie Maes. 2009. SixthSense: Integrating information and the real world. In *Science & Technology Proceedings, 8th IEEE International Symposium on Mixed and Augmented Reality 2009, ISMAR 2009, Orlando, Florida, USA, October 19-22, 2009.* IEEE Computer Society. https://doi.org/10.1109/ISMAR.2009.5336511

[38] Alexander Masurovsky, Paul Chojecki, Detlef Runde, Mustafa Lafci, David Przewozny, and Michael Gaebler. 2020. Controller-Free Hand Tracking for Grab-and-Place Tasks in Immersive Virtual Reality: Design Elements and Their Empirical Study. *Multimodal Technologies and Interaction* 4, 4 (2020). https://doi.org/10.3390/mti4040091

[39] Rory McGloin, Kirstie Farrar, and Marina Krcmar. 2013. Video Games, Immersion, and Cognitive Aggression: Does the Controller Matter? *Media Psychology* 16 (01 2013), 65. https://doi.org/10.1080/15213269.2012.752428

[40] Meta. 2019. Introducing hand tracking on oculus quest-bringing your real hands into vr. https://www.oculus.com/blog/introducing-hand-tracking-on-oculus-quest-bringing-your-real-hands-into-vr/ Retrieved January 01, 2022.

[41] Microsoft. 2019. Hand menu. https://docs.microsoft.com/en-us/windows/mixed-reality/design/hand-menu Retrieved January 01, 2022.

[42] Microsoft. 2019. Introducing instinctual interactions. https://docs.microsoft.com/en-us/windows/mixed-reality/design/interaction-fundamentals Retrieved January 01, 2022.

[43] Microsoft. 2021. Hand tracking - mixed reality toolkit. https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/features/input/hand-tracking Retrieved January 01, 2022.

[44] Adiyan Mujibiya, Xiang Cao, Desney S. Tan, Dan Morris, Shwetak N. Patel, and Jun Rekimoto. 2013. The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. In *The ACM International Conference on Interactive Tabletops and Surfaces, ITS '13, St Andrews, United Kingdom - October 06 - 09, 2013.* ACM, 189–198. https://doi.org/10.1145/2512349.2512821

[45] Nick Babich. 2020. UX Design Principles for Augmented Reality. https://xd.adobe.com/ideas/principles/emerging-technology/ux-design-principles-for-augmented-reality/ Retrieved January 01, 2022.

[46] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Conference on Human Factors in Computing Systems, CHI 1990, Seattle, WA, USA, April 1-5, 1990, Proceedings*. ACM, 249–256. https://doi.org/10.1145/97243.97281

[47] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

[48] Ken Pfeuffer, Lukas Mecke, Sarah Delgado Rodriguez, Mariam Hassib, Hannah Maier, and Florian Alt. 2020. Empirical Evaluation of Gaze-enhanced Menus in Virtual Reality. In *VRST '20: 26th ACM Symposium on Virtual Reality Software and Technology, Virtual Event, Canada, November 1-4, 2020*. ACM, 20:1–20:11. https://doi.org/10.1145/3385956.3418962

[49] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology, UIST 1996, Seattle, WA, USA, November 6-8, 1996*. ACM, 79–80. https://doi.org/10.1145/237091.237102

[50] Jing Qian, Jiaju Ma, Xiangyu Li, Benjamin Attal, Haoming Lai, James Tompkin, John F. Hughes, and Jeff Huang. 2019. Portal-ble: Intuitive free-hand manipulation in unbounded smartphone-based augmented reality. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (Oct 2019). https://doi.org/10.1145/3332165.3347904

[51] Zulqarnain Rashid, Joan Melià-Seguí, Rafael Pous, and Enric Peig. 2017. Using Augmented Reality and Internet of Things to improve accessibility of people with motor disabilities in the context of Smart Cities. *Future Gener. Comput. Syst.* 76 (2017), 248–261. https://doi.org/10.1016/j.future.2016.11.030

[52] Sofia Seinfeld, Tiare M. Feuchtner, Antonella Maselli, and Jörg Müller. 2021. User Representations in Human-Computer Interaction. *Hum. Comput. Interact.* 36, 5-6 (2021), 400–438. https://doi.org/10.1080/07370024.2020.1724790

[53] Mohamed Soliman, Franziska Mueller, Lena Hegemann, Joan Sol Roo, Christian Theobalt, and Jürgen Steimle. 2018. FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (Tokyo, Japan) *(ISS '18)*. Association for Computing Machinery, New York, NY, USA, 177–187. https://doi.org/10.1145/3279778.3279799

[54] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. 2017. WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017*. ACM, 3891–3902. https://doi.org/10.1145/3025453.3026005

[55] Jürgen Steimle, Joanna Bergstrom-Lehtovirta, Martin Weigel, Aditya Shekhar Nittala, Sebastian Boring, Alex Olwal, and Kasper Hornbæk. 2017. On-skin interaction using body landmarks. *Computer* 50, 10 (2017), 19–27.

[56] Jürgen Steimle, Joanna Bergstrom-Lehtovirta, Martin Weigel, Aditya Shekhar Nittala, Sebastian Boring, Alex Olwal, and Kasper Hornbæk. 2017. On-Skin Interaction Using Body Landmarks. *Computer* 50, 10 (2017), 19–27. https://doi.org/10.1109/MC.2017.3641636

[57] Frank Steinicke, Timo Ropinski, and Klaus Hinrichs. 2006. Object selection in virtual environments using an improved virtual pointer metaphor. In *Computer vision and graphics*. Springer, 320–326.

[58] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental Analysis of Barehand Mid-air Mode-Switching Techniques in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*. ACM, 196. https://doi.org/10.1145/3290605.3300426

[59] Amato Tsuji, Keita Ushida, and Qiu Chen. 2018. Real Time Animation of 3D Models with Finger Plays and Hand Shadow. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces, ISS 2018, Tokyo, Japan, November 25-28, 2018*. ACM, 441–444. https://doi.org/10.1145/3279778.3279918

[60] Tyler Wilson. 2017. The principles of good UX for Augmented Reality. https://uxdesign.cc/the-principles-of-good-user-experience-design-for-augmented-reality-d8e22777aabd Retrieved January 01, 2022.

[61] Ultraleap. [n.d.]. Leap motion controller. https://www.ultraleap.com/product/leap-motion-controller/

[62] Ultraleap. 2021. Designing for Hand Tracking. https://docs.ultraleap.com/xr-guidelines/designing-hand-tracking/ Retrieved January 01, 2022.

[63] Lode Vanacken, Tovi Grossman, and Karin Coninx. 2007. Exploring the Effects of Environment Density and Target Visibility on Object Selection in 3D Virtual Environments. In *IEEE Symposium on 3D User Interfaces, 3DUI 2007, Charlotte, North Carolina, USA, 10-11 March, 2007*. IEEE Computer Society, 27. https://doi.org/10.1109/3DUI.2007.340783

[64] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*. ACM, 42. https://doi.org/10.1145/3173574.3173616

[65] Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, and Yuanchun Shi. 2018. VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*. ACM, 78. https://doi.org/10.1145/3173574.3173652

[66] Zhixin Yan, Robert W. Lindeman, and Arindam Dey. 2016. Let your fingers do the walking: A unified approach for efficient short-, medium-, and long-distance travel in VR. In *2016 IEEE Symposium on 3D User Interfaces, 3DUI 2016, Greenville, SC, USA, March 19-20, 2016*. IEEE Computer Society, 27–30. https://doi.org/10.1109/3DUI.2016.7460027

[67] Shumin Zhai, William Buxton, and Paul Milgram. 1994. The "Silk Cursor": investigating transparency for 3D target acquisition. In *Conference on Human Factors in Computing Systems, CHI 1994, Boston, Massachusetts, USA, April 24-28, 1994, Proceedings*. ACM, 459–464. https://doi.org/10.1145/191666.191822

[68] Yang Zhang, Wolf Kienzle, Yanjun Ma, Shiu S. Ng, Hrvoje Benko, and Chris Harrison. 2019. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, UIST 2019, New Orleans, LA, USA, October 20-23, 2019*. ACM, 1151–1159. https://doi.org/10.1145/3332165.3347869

[69] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. 2019. SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*. ACM, 111. https://doi.org/10.1145/3290605.3300341

[70] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner, and Shiri Azenkot. 2019. Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, UIST 2019, New Orleans, LA, USA, October 20-23, 2019*. ACM, 387–402. https://doi.org/10.1145/3332165.3347906