Title
Designing Augmented Reality and Virtual Reality Interfaces around Hand Expressivity

Author
Pei, Siyou

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Designing Augmented Reality and Virtual Reality Interfaces around Hand Expressivity

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Electrical and Computer Engineering

by

Siyou Pei

2021

ABSTRACT OF THE THESIS

Designing Augmented Reality and Virtual Reality Interfaces
around Hand Expressivity

by

Siyou Pei
Master of Science in Electrical and Computer Engineering
University of California, Los Angeles, 2021
Professor Yang Zhang, Chair

Augmented Reality (AR) and Virtual Reality (VR) create exciting new opportunities for people to interact with computing resources and information. Less so exciting is the need for holding hand controllers, which limits applications that demand expressive readily available interactions. In response, prior research has investigated freehand AR/VR input by transforming the user body into an interaction medium. While prior work focuses on having users' hands grasp virtual objects, we propose a new interaction technique to have users' hands become virtual objects through imitation, for example, having a thumb-up hand pose to imitate a joystick. We have created a wide array of interaction designs around this idea to demonstrate its applicability in object retrieval and interactivity. Collectively, we call these interaction designs Hand Interfaces. With a series of user studies comparing Hand Interfaces against various baseline techniques, we have collected quantitative and qualitative feedback which indicated that Hand Interfaces is effective, expressive, and fun to use.

The thesis of Siyou Pei is approved.

Ankur M. Mehta

Xiang Chen

Yang Zhang, Committee Chair

University of California, Los Angeles

2021

TABLE OF CONTENTS

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Thesis Outline 3

2 Literature Review **5**

2.1 AR/VR Interaction Design Principles and Heuristics 5

2.2 Free-hand AR/VR Interactions 6

2.3 Leveraging the User Body as AR/VR Interaction Medium 7

3 Interaction Design **10**

3.1 Heuristic Idea Generation 11

3.2 Criterion of Good Designs 11

4 Implementation **15**

4.1 Hardware 15

4.2 Software 15

5 Study of Object Retrieval **19**

5.1 Baseline Designs 19

5.2 Evaluation Configurations 20

5.2.1 Participants 21

5.2.2 Procedure 21

5.3 Evaluation Results 22

5.3.1 Quantitative Feedback 22
5.3.2 Qualitative Feedback — Real-world Experience on the Perception of Realism 24
5.3.3 Qualitative Feedback — Visuals on the Perception of Realism 26
5.3.4 Qualitative Feedback — Ease of Recollection 27
5.3.5 Qualitative Feedback — Comfort 28

6 Study of Interactive Control 30
6.1 Baseline Designs 30
6.2 Evaluation Configurations 31
6.2.1 Procedures 31
6.3 Evaluation Results 32
6.3.1 Quantitative Feedback 32
6.3.2 Qualitative Feedback — Real World Experience on the Perception of Realism 35
6.3.3 Qualitative Feedback — Tactile Feedback 36
6.3.4 Qualitative Feedback — Ease of Recollection 36
6.3.5 Qualitative Feedback — Constraints from Hand Kinematics 37
6.3.6 Qualitative Feedback — Miscellaneous Insights 38

7 Example Use Cases 40
7.1 Education 40
7.2 Entertainment 40
7.3 Ubiquitous Computing 41

8 Discussion and Conclusion **43**

A Design-wise Score percentages Diagrams **46**

A.1 Retrieval Evaluation of Drop-down Menu (the First Baseline) 46

A.2 Retrieval Evaluation of VirtualGrasp (the Second Baseline) 49

A.3 Retrieval Evaluation of Hand Interfaces 51

A.4 Interactivity Evaluation of Fist Gesture (the First Baseline) 54

A.5 Interactivity Evaluation of Virtual Manipulation (the Second Baseline) . . . 57

A.6 Interactivity Evaluation of Hand Interfaces 59

LIST OF FIGURES

1.1 Hand Interfaces allows users to imitate a wide range of objects which we perceive as AR/VR interfaces, for expressive readily available interactions. The figure demonstrates how users imitate a joystick, pair of binoculars, book, camera, fishing rod, flower, fork, magnet, lever, kalimba, hourglass, pigeon, heart, inflator, globe, ladle, scissors, rake, pen, multi-meter probe, cup, water blaster, wrench, wand, trumpet, toggle switch, spray can, or stapler. 3

3.1 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. 12

4.1 The detection pipeline of Hand Interfaces. Once the headset cameras detect hands, the algorithm will track if a user is doing a designed gesture by matching hand keypoints with gesture templates. Corresponding virtual objects will be rendered on hands when predefined gestures are detected. 16

5.1 In this study we explored three techniques for the object retrieval scenarios, including two baselines and hand interfaces. The figure shows what retrieving a joystick looks like with each technique, from left to right: Drop-down Menu, VirtualGrasp, and Hand Interfaces. 20

5.2 Retrieval Evaluation of Drop-down Menu, VirtualGrasp, and Hand Interfaces. The first (Blue), second (Red) and third (yellow) bar in each bar set indicates averaged scores 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 degree of significance between HI and baseline techniques. From the figure, we learn HI received significantly higher score than DM in fidelity during retrieval. 23

5.3 Score percentages of VirtualGrasp (the second baseline) and Hand Interfaces per Interface on Fidelity. Seven colors denote 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 in horizontal axis while interface-wise user preferences are visible across each row. From the figure, we know that in HI, Globe, Scissors, Binoculars that involved gestures people commonly make in reality received the most positive feedback. 25

6.1 In this study we explored three techniques for the interactive control scenarios, including two baselines and hand interfaces. The figure shows what interacting with a joystick looks like with each technique, from left to right: Fist Gesture, Virtual Manipulation, and Hand Interfaces. 31

6.2 Interaction Evaluation of Fist Gesture, Virtual Manipulation, and Hand Interfaces. The first (Blue), second (Red) and third (yellow) bar in each bar set indicates averaged scores of FG, VM, and HI techniques, with 1 being "strongly disagree" and 7 being "strongly agree" on a 7-point Likert scale. The number of asterisks denotes degree of significance between HI and baseline techniques. From the figure, we learn HI received higher score than FM in fidelity during interactive control with relative significance. 33

6.3 Score percentages of Fist Gesture (the first baseline), and Hand Interfaces per Interface on Freedom of Movement. The figure indicated a clear preference for the FG Fishing rod over that of the HI with respect to freedom of movement. . 34

7.1 Hand Interfaces allows users to easily use various tools in an educational setting. Specifically, *Multi-meter probes* are used to measure voltages (A), the *Pen* is used for writing notes (B), and the *Camera* (C) is used for taking a photo of the current scene. 41

7.2 In a magic fighting game, a user uses the hand-imitated *Wand* to cast spells (A), the *Mug* to use healing potion (B), and the *Book* to level up (C). 41

7.3 A user uses Hand Interfaces in concert with an AR device (A, B) to quickly and easily control a smart lighting system. Specifically, the *Toggle switch* is used to turn on/off the light (C), the *Joystick* controls its pan/tilt (D), and a fist-imitated *Color palette* is used to set its color (E). 42

A.1 Retrieval Evaluation of Drop-down Menu per Interface on Fidelity 46

A.2 Retrieval Evaluation of Drop-down Menu per Interface on Movement 47

A.3 Retrieval Evaluation of Drop-down Menu per Interface on Swiftness 47

A.4 Retrieval Evaluation of Drop-down Menu per Interface on Comfort 48

A.5 Retrieval Evaluation of Drop-down Menu per Interface on Recollection 48

A.6 Retrieval Evaluation of VirtualGrasp per Interface on Fidelity 49

A.7 Retrieval Evaluation of VirtualGrasp per Interface on Movement 49

A.8 Retrieval Evaluation of VirtualGrasp per Interface on Swiftness 50

A.9 Retrieval Evaluation of VirtualGrasp per Interface on Comfort 50

A.10 Retrieval Evaluation of VirtualGrasp per Interface on Recollection 51

A.11 Retrieval Evaluation of Hand Interfaces per Interface on Fidelity 51

A.12 Retrieval Evaluation of Hand Interfaces per Interface on Movement 52
A.13 Retrieval Evaluation of Hand Interfaces per Interface on Swiftness 52
A.14 Retrieval Evaluation of Hand Interfaces per Interface on Comfort 53
A.15 Retrieval Evaluation of Hand Interfaces per Interface on Recollection 53
A.16 Interactivity Evaluation of Fist Gesture per Interface on Fidelity 54
A.17 Interactivity Evaluation of Fist Gesture per Interface on Movement 54
A.18 Interactivity Evaluation of Fist Gesture per Interface on Swiftness 55
A.19 Interactivity Evaluation of Fist Gesture per Interface on Comfort 55
A.20 Interactivity Evaluation of Fist Gesture per Interface on Recollection 56
A.21 Interactivity Evaluation of Virtual Manipulation per Interface on Fidelity 57
A.22 Interactivity Evaluation of Virtual Manipulation per Interface on Movement . . 57
A.23 Interactivity Evaluation of Virtual Manipulation per Interface on Swiftness . . . 58
A.24 Interactivity Evaluation of Virtual Manipulation per Interface on Comfort . . . 58
A.25 Interactivity Evaluation of Virtual Manipulation per Interface on Recollection . 59
A.26 Interactivity Evaluation of Hand Interfaces per Interface on Fidelity 59
A.27 Interactivity Evaluation of Hand Interfaces per Interface on Movement 60
A.28 Interactivity Evaluation of Hand Interfaces per Interface on Swiftness 60
A.29 Interactivity Evaluation of Hand Interfaces per Interface on Comfort 61
A.30 Interactivity Evaluation of Hand Interfaces per Interface on Recollection 61

LIST OF TABLES

2.1 Design table that highlights the novelty of Hand Interfaces among prior literature in terms of interface location and modality. Hand Interfaces opens up a new window for on-body interface in 3D context. 9

ACKNOWLEDGMENTS

I would like to first express my deepest appreciation to thesis Committee Chair, Prof. Yang Zhang of the Electrical and Computer Engineering (ECE) Department at University of California, Los Angeles (UCLA). His passion to steer me in the right direction turned me into a more mature researcher. His unwavering support with profound belief made me a better person. I would also like to extend my deepest gratitude to Prof. Xiang 'Anthony' Chen, and Prof. Ankur M. Mehta of the ECE Department at UCLA who serve as members of my master thesis committee for their active participation and invaluable insights.

I must also thank Prof. Mani Srivastava of the ECE Department at UCLA, who guided me with patience during my first year. Without his generous support and constructive advice, I would not have acquired such practical experience. I'm also extremely grateful to Deeona Columbia, Office of Graduate Student Affairs' Manager of the ECE Department at UCLA for offering unparalleled support and a great amount of assistance, without whom I would never have accomplished this.

I also wish to thank my collaborators. The thesis is a modified version of an accepted manuscript [41] at CHI 2022. It is not yet published when I submit the thesis. I had great pleasure of working with Alexander Chen, an excellent senior undergraduate of the ECE Department at UCLA, a great collaborator and friend, with strong capabilities and responsibility. I also gratefully acknowledge the research efforts of Jaewook Lee, a senior undergraduate at the University of Illinois at Urbana-Champaign.

Thanks should also go to Qian Gao, Weixin Liang, Ziqi Wang, Ruolin Wang, Pengrui, Hengjie Yang, Weiqi Liu and Bo-Jhang Ho for offering helpful suggestions as researchers and/or support as friends. I also wish to thank Quentin Valembois, Dilmer Valecillos and Antony Vitillo for their contribution in great AR/VR content creation. Special thanks to the authors of VirtualGrasp [56], Yukang Yan, et al. for inspiring this thread of work in designing AR/VR interfaces around hands' expressivity. Many thanks to all research participants for completing user studies seriously.

Finally, I must express my very profound gratitude to my parents and to my partner Weixin Liang for providing me with unfailing support and continuous encouragement throughout my years of study. Their curiosity about my research has cured me again and again. This accomplishment would not have been possible without their unconditional love. Thank you, your help can never be overestimated.

CHAPTER 1

Introduction

1.1 Background and Motivation

Augmented Reality (AR) renders a digital display layer upon the physical world settings, enabling people to see 3D virtual objects in real environments. Virtual Reality (VR) creates an immersive virtual 3D world while occluding the view of the physical world. AR/VR has shown great promise in education [27], accessibility [45], and health care [20], and has a rapidly growing user base that makes many predict that it will be the next ubiquitous device (after smartphones) that fundamentally changes human-computer interaction. With AR/VR headsets becoming increasingly popular, there has been a significant amount of research on techniques that allow users to easily and naturally manipulate AR/VR content. 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. Prior work has investigated interaction techniques designed around user hands [56, 26, 14, 19, 36, 51].

Both this and prior work look into the design space of hand-centric interactions for AR/VR. However, prior work either used hands as 2D surfaces for touch interactions (e.g., Finger Input [46], ActiTouch [58], It's a Wrap [14], SkinMarks [48]), or discrete controllers for mode switching (e.g., Open Palm Menu [11], Surale et al. [51]). Little has been done considering user hands as expressive 3D structures to host interactions. One of the few prior systems that leverage the hands' 3D expressivity is VirtualGrasp [56] which lets users retrieve a virtual object by performing hand poses as if they were grasping the object. Hand Interfaces directly extends 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 expressivity of human hands has long been used in our daily lives but is under-explored in AR/VR applications. In this research, we hope to demonstrate the potential of this idea in implementing expressive retrieval and interactivity, and evaluate the pros and cons of this idea with user studies. Figure 1.1 shows 28 example designs of Hand Interfaces. 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 (Fig. 1.1a). Similarly, extended fingers can imitate Kalimba keys (Fig. 1.1j), while users can play simple music by tapping keys with another hand. Another instance is the joint of an index finger alone can emulate a toggle switch (Fig. 1.1z). All these examples utilize the expressivity of users' hands and provide a rich user experience in AR/VR.

Another significance of this research lies in the evaluation and quantification of the benefit of hand-centric user interfaces. As we mentioned earlier, there has been increasing research momentum on this topic but many ideas have not been supported with real users. This lack of investigation makes it difficult to assess its 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 idea, we have also evaluated it with a wide array of designs (i.e., 11 controllers) in two interaction scenarios (i.e., object retrieval and interactive control). 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.

Figure 1.1: Hand Interfaces allows users to imitate a wide range of objects which we perceive as AR/VR interfaces, for expressive readily available interactions. The figure demonstrates how users imitate a joystick, pair of binoculars, book, camera, fishing rod, flower, fork, magnet, lever, kalimba, hourglass, pigeon, heart, inflator, globe, ladle, scissors, rake, pen, multi-meter probe, cup, water blaster, wrench, wand, trumpet, toggle switch, spray can, or stapler.

1.2 Thesis Outline

In this research, we first systematically reviewed prior work and guidelines on using bare hands as expressive controls for AR/VR in Chapter 2. We then designed a wide array of interaction techniques based on Hand Interfaces in Chapter 3 and built proof-of-concept detection pipelines with an Oculus Quest headset and its hand tracking feature[4], which is illustrated in Chapter 4. In Chapter 5 and 6, we evaluated Hand Interfaces with 11 distinct interface designs with respect to object retrieval and interactive control in a series of user studies, where both qualitative and quantitative feedback was gathered from 17 participants. In Chapter 7, we presented multiple example use cases of Hand Interfaces. Overall, the results indicated that Hand Interfaces is effective, expressive, and fun to use. Finally, we discussed the advantages and limitation of the work, and concluded it with future work in Chapter 8. Design-wise score percentage diagrams of all metrics and all interaction techniques are attached in the Appendix.

CHAPTER 2

Literature Review

In this section, we first talk about design principles and heuristics from prior work on creating effective AR/VR interactions, which we took inspirations from when creating Hand Interfaces. Then, we review the three key bodies of research in AR/VR that are closely related to Hand Interfaces.

2.1 AR/VR Interaction Design Principles and Heuristics

Design principles and heuristics help evaluate the usability of interaction techniques and guide their designs. 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 [40]. 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 [8]. Microsoft shared their design philosophy of AR interaction utilizing on hands and arms [3, 5]. The increasing availability of commercially available AR/VR hardware has lowered the barrier which has led to a rapidly growing user community of AR/VR, and shared design recommendations [6, 2, 1, 7].

In the research domain, Dv̈nser et al. [17] 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 and feedback, and error tolerance. More recently, Endsley et al. [18] generated eight heuristics with an iterative process working closely with experts and designers. These 8 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. Finally, beyond these general-purpose guidelines, researchers have also proposed guidelines for specific platforms (e.g., mobile computing [31]), applications (e.g., education [28]), and user groups (e.g., people with low vision [60, 59]).

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 [36, 37]. 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 [42] and voice [22], and 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 commercially available products (e.g., Leap Motion [53], HoloLens [29], and Oculus Quest [4]) that use computer vision to track hands in close range, which is ideal for AR/VR since AR/VR devices 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 [30], bio-acoustic vibrations [25], active ultrasonic sensing [39] and skin-mediated radio frequency [58].

With hand tracking, prior research has designed interaction techniques around the expressivity of users' hands. For example, Open Palm Menu [11] 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 [44] proposed sensing and interaction techniques for users to grasp and manipulate virtual objects in smartphone-based augmented reality. Surale et al. [51] explored seven bare hand interaction techniques for mode-switching tasks in VR. Similarly, Masurovsky et al. [36] investigated the performance of controller-free hand interactions in grab-and-place tasks. Both research yielded quantitative and qualitative results that ground the benefits of free-hand interactions and provide footholds for future research in this domain. 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 [16]. In another example, PinchType [19] enables users to type with thumb to fingertip pinches. Closely related to Hand Interfaces is VirtualGrasp [56], which 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. This ingenious leverage 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 leverages the user body as AR/VR interaction media which user interfaces can be linked to or reside on. First, the user body (e.g., hands and arms) can serve as spatial references to graphical menus [32] and user interactions [47] to facilitate natural and precise input. Specifically, Lediaeva et al. [32] investigated methods to render graphical interfaces around user hands and arms for AR/VR inputs. WatchSense [47] uses fingers touching on the arm as reference points that open up a rich set of finger gestures. Yan et al. [55] 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 [13, 49].

Researchers have also used body surfaces to host conventional GUIs in 2D. For example, SixthSense [35], 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 [14]). There also exists a large body of research on on-body interactions not specifically designed for AR/VR applications but can be easily adapted. For a comprehensive review of literature, we recommend this survey [12]. In all prior systems, users need to straighten their palms and fingers to make planar surfaces, best at hosting displays and touch inputs for GUIs. In contrast, Hand Interfaces does not suppress the 3D expressivity of hands, but instead leverages 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. Tsuji et al. [52] proposed a method that allows users to animate 3D models with finger play and hand shadow. For example, users can imitate opening and closing a crocodile's jaw by opening and closing their palms. In a more general-purpose UI design, DigiGlo [15] 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 split attention avoidance. Finally, prior research has demonstrated transforming the user body into actuation mechanisms which we believe constitute an important aspect of 3D interfaces. For example, Lopes et al.

		Interface Modality	
		0D/1D/2D/2.5D	3D
Interface Location	Off-body	GUIs decoupled from users' body (e.g., pointing) [4, 11, 16, 19, 32, 50, 54]	3D virtual objects that do not share voxels with users' body (e.g., direct manipulation) [9, 26, 33, 34, 36, 42, 43, 55]
	On-body	GUIs that reside on users' body (e.g., digital glove) [14, 15, 24, 25, 35, 57]	3D virtual objects that share voxels with users' body **Hand Interfaces**

Table 2.1: Design table that highlights the novelty of Hand Interfaces among prior literature in terms of interface location and modality. Hand Interfaces opens up a new window for on-body interface in 3D context.

demonstrated adding Electrical Muscle Stimulation (EMS) to heavy objects (e.g., walls) in virtual reality [33], and using force feedback in a wide variety of mixed reality scenarios to enhance user experience [34]. 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.

As we discussed above, there has been much effort on 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., new type of mouse), or on on-body interfaces that wrap around the user body, like digital skin or gloves. Hand Interfaces first looked into ways to transform a user's body into virtual objects, which we think of as 3D interfaces to get information from or actuate the virtual world. As we will demonstrate with a rich set of examples later in this the paper, we believe our interaction techniques could benefit future AR/VR in a wide spectrum of scenarios. Table 2.1 summarizes the novelty of Hand Interfaces among prior literature on free-hand interactions for AR and VR.

CHAPTER 3

Interaction Design

Overall, we believe in several innate advantages of Hand Interfaces, around which we designed our interaction techniques. First, like other free-hand interactions, Hand Interfaces are readily available to users and thus is 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 offers tactile feedback by nature due to skin contacts. Finally, Hand Interfaces leverages proprioception that improves the precision of 3D manipulations. Compared with conventional free-hand interactions, Hand Interfaces can be more expressive in some cases. For example, techniques that leverage grasping postures to allow users to directly "grasp" virtual objects yield ambiguities in retrieving objects when different objects might have similar grasping postures. This is not uncommon given that common practices improve affordance of everyday objects by adopting universal industry designs. For example, a virtual stapler might have users grasping it in a way same as how users grasp a wand, a fork, a billiard cue, or anything that features a pole-like user interface. In our design process, we aimed to make the most of these advantages 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, the process of which we discuss in this section.

3.1 Heuristic Idea Generation

We conducted concept-driven brainstorming with all researchers in this project. Researchers were asked to come up with designs that use hands to imitate objects as user interfaces. Early ideas involved mostly digital interfaces inspired by prior work and conventional computer platforms. Such designs include, for example, 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. Then, ideas generated by researchers quickly evolved into a wide array of 3D objects. We started with common controllers in the real world. Examples include a joystick imitated by a thumbs-up gesture (with the thumb "becoming"the stick), a mouse (with two fingers as the left and right button), and a toggle switch (imitated by the index finger joint). Finally, ideas from researchers opened up to a wider array of objects, which were less of interfaces by design but more-so props users could utilize in AR/VR environments. Examples include tools such as a mug, hourglass, spray can, microphone, pair of scissors, wrench, fork, ladle, lever, binoculars, fishing rod, and pump; musical instruments such as a trumpet, bongo, and kalimba; educational and entertainment props such as a globe, multi-meter, color palette, and wand.

3.2 Criterion of Good Designs

In the course of designing interactions, we leveraged several design considerations in the selection process which we list as follows. We use our design process of *Spray can* (Figure 3.1) as an illustrative example.

1. The first consideration was **shape similarity**. We first removed design ideas with hand poses that least resembled the target objects. As we later found in the study, considering shape similarity also contributed to users' perception of realism.

2. Another consideration we adopted was **kinematic similarity**. In this consideration,

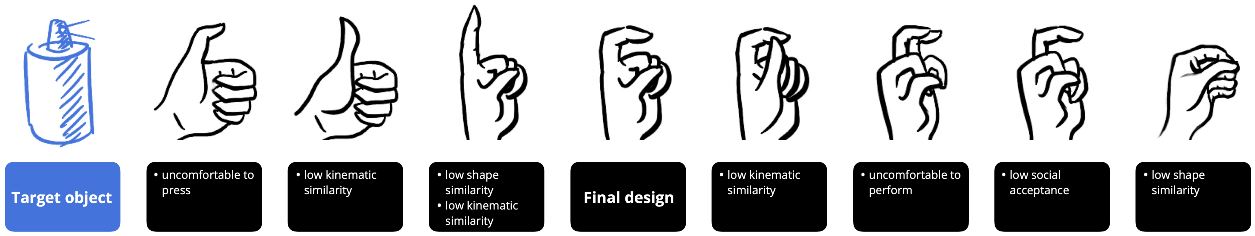

Figure 3.1: Designs that were conceived for the *Spray can* during brainstorming. We considered shape similarity, kinematic similarity, comfort, and social acceptance to pick out the best design, which in this case is the fourth gesture.

we estimated how similar the dynamic characteristics (e.g., degree of freedom) of hands and objects were. In the *Spray can* example, a spray nozzle is meant to be pressed and thus we removed designs that provided no "pressing-in" mechanism. In practice, we found that considering kinematic similarity was critical to the comfort of Hand Interfaces designs.

3. We also considered **comfort**. Specifically, we avoided hand poses that were difficult to perform or uncomfortable to maintain. We also removed designs that felt uncomfortable during interactions when the manipulating hand pressed on the imitating hand.

4. Finally, we considered **social acceptance**. Hand poses should be socially acceptable as AR/VR has a wide range of applications involving multiple users or in public spaces. Therefore, having socially acceptable designs is a prerequisite to success. In the *Spray can* example, we removed the design with a curled middle finger for the consideration of social acceptance.

Eventually, our design process yielded 28 Hand Interfaces shown in Figure 1.1. On a high level, these designs can be categorized by number of hands involved. 22 designs involved only one hand and the rest involved two hands. Specifically, 10 out of 22 one-hand designs required manipulations from the other hand. Most of these designs were self-revealing. For example, to use the *Ladle* and the *Fork*, users would perform the same interaction as they would in reality. Similarly, designs that involved both hands were intuitive as well. For example, the *Binoculars* could be raised up close to users' eyes to transition their view from normal to long-range. A tad more complicated were single-hand designs that involved the other hand for manipulations, which we describe below:

- *Joystick* interface (Fig. 1.1a) required users to grasp the thumb of the imitating hand (the stick of the joystick) and move it around to control something, e.g. the direction of a game character.

- *Fishing rod* (Fig. 1.1e) was imitated by pointing the thumb horizontally to one side as the fishing 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* (Fig. 1.1i) 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* (Fig. 1.1j), 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 virtual piano keys of the kalimba and create a simple melody.

- *Inflator* (Fig. 1.1n), 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* (Fig. 1.1o) was basically a sphere 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* (Fig. 1.1y) 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* (Fig. 1.1z) was a switch rendered on the first joint of the index finger of the imitating hand. A user could then click on the joint to perform interactions, e.g., turn on/off the switch to toggle virtual lights.

- *Spray can* (Fig. 1.1α), was represented by a hook-like hand gesture where the index finger imitated the nozzle and the other fingers acted like 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 1.1β) 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 a comprehensive demonstration of these Hand Interfaces in action, please refer to the Video Figure. Inspired by VirtualGrasp [56], we found Hand Interfaces also useful in object retrieval. Specifically, users can perform a hand pose to retrieve the corresponding virtual object for further interactivity. We have implemented a detection pipeline using a commercially available VR headset (i.e., Oculus Quest) to demonstrate the feasibility of Hand Interfaces. We will talk about our implementation of Hand Interfaces in detail in next section.

CHAPTER 4

Implementation

Hand Interfaces was built on commercially available hardware with custom designed detection algorithms based on existing hand tracking APIs. In this section we describe both 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 a RTX 3070 8G GPU) using a USB 3.0 Type-C cable. Four built-in cameras on the Quest enables hand tracking in real-time. We used two Oculus Touch Controllers to implement baseline designs in our user studies.

4.2 Software

Hand Interfaces software was built on Unity platform (2020.3.7f1 ver) and is summarized in Figure 4.1. First, the system initializes at the state of "no hand" 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 ~ 35 FPS. 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 be either designed by authors of the application in advance, or defined by users during

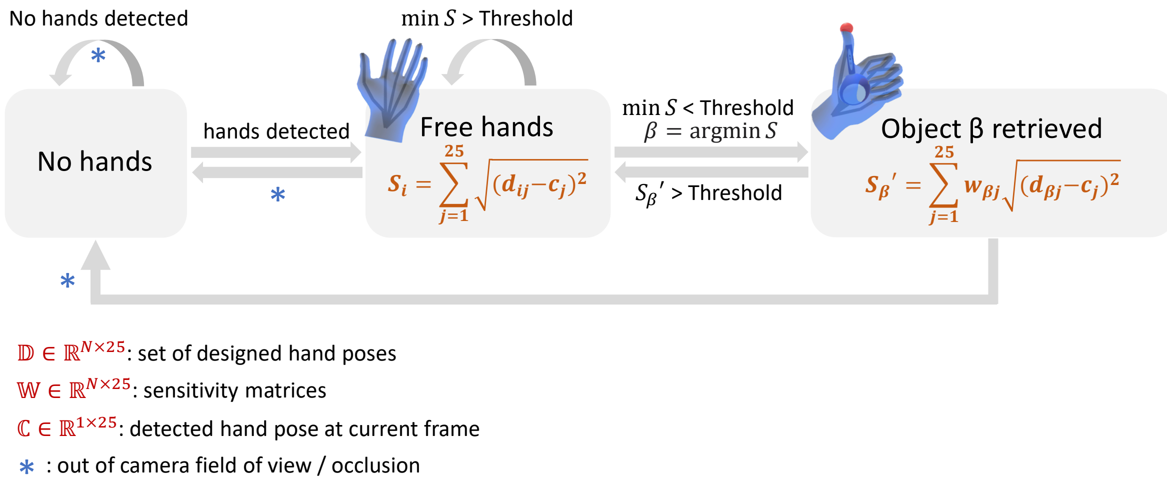

Figure 4.1: The detection pipeline of Hand Interfaces. Once the headset cameras detect hands, the algorithm will track if a user is doing a designed gesture by matching hand keypoints with gesture templates. Corresponding virtual objects will be rendered on hands when predefined gestures are detected.

run time. The N in Figure 4.1 denotes the number of gesture templates in the dataset. The difference score S_i is calculated by a sum of Euclidean distances per pair of 25 joints each hand between the current gesture and the i^{th} gesture. Therefore, the more similar the current gesture is to a gesture template, the smaller the difference score is.

Our algorithms keep track of the minimal difference score in the gesture map. If $minS$ is higher than a pre-defined threshold, the software stays at the "free hand" state. Once $minS$ falls below the threshold, the software transitions to the "object retrieved" state as the algorithms determine object/gesture β is retrieved/detected where $\beta = argminS$.

When the current state is "object retrieved", users can interact with the current virtual object. We introduced a weighted distance score S'_β using sensitivity matrices W_β, each of which is tuned for specific Hand Interfaces designs to make the software less likely to dismiss the object while users are interacting with it. Specifically, the sensitivity matrices assign lower weights for key points that are supposed to move around during interactions, and higher weights for static key points. If S'_β is greater than the threshold, the software dismisses the object and transitions back to the "free hand" state.

Once objects are retrieved, the 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 categorize our detection methods by the interaction designs and describe each with more details:

- For designs that rely on discrete hand poses users need to perform (e.g *Scissors*). We use the same software described above. For example, once the index and middle fingers are detected as parallel and touching in the *Scissors* hand pose, the action of cutting is performed.

- For designs that involve proximity-based interactions (e.g., *Binoculars*), we check distances between the anchoring points on the imitating hands and those on objects in the environment. For example, the *Binoculars* design presents users a long-range view when the *Binoculars* are raised up close to users' eyes.

- For designs that users click (e.g., *Globe*, *Toggle switch*), we continuously monitor the distance between joints of the manipulating hand and joints of the imitating hand. We detect clicks by looking for patterns of "approach, touch, and no touch". We hard coded the distance threshold to distinguish between touch and no touch to 7 mm.

- For designs that feature handle-like interactions (e.g., *Joystick* and *Fishing rod*), we again monitor distances between touching joints and touched joints, in addition to which we also detect “approach, grasp, and movement”. In other words, we track if the interacting hand performs manipulations on top of tracking clicks. There are designs that rely on more complicated manipulating hand poses such as *Inflator* and *Spray can*, which involve more complicated detection heuristics consisting of simpler ones used in previous designs.

Additionally, we fine tuned thresholds for the difference score and sensitivity matrices of each Hand Interfaces design for a precise detection. We implemented a hysteresis buffer of 200 milliseconds of retrieval results to improve robustness. We also cached latest frames of hand tracking which we used to replace current frames to remove jitters in Oculus hand tracking, which could happen during brief occlusions.

CHAPTER 5

Study of Object Retrieval

Object retrieval is an important building block for AR/VR interactions, as many previous AR/VR systems aimed to create new retrieval interaction modalities [55, 57, 54, 43, 50, 23]. In response, 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. These objects can be found on Figure 5.1 left. Details of hand poses to retrieve these objects can be found on Figure 1.1.

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 miniaturised objects on a flat 2D plane in a virtual 3D environment. This is similar to most GUIs in people's daily use of computers (Figure 5.1 left). The models rotated at a speed of 18 degrees/second 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 [56] (Figure 5.1 middle), in which participants used their hands to grasp objects at positions and with poses suggested

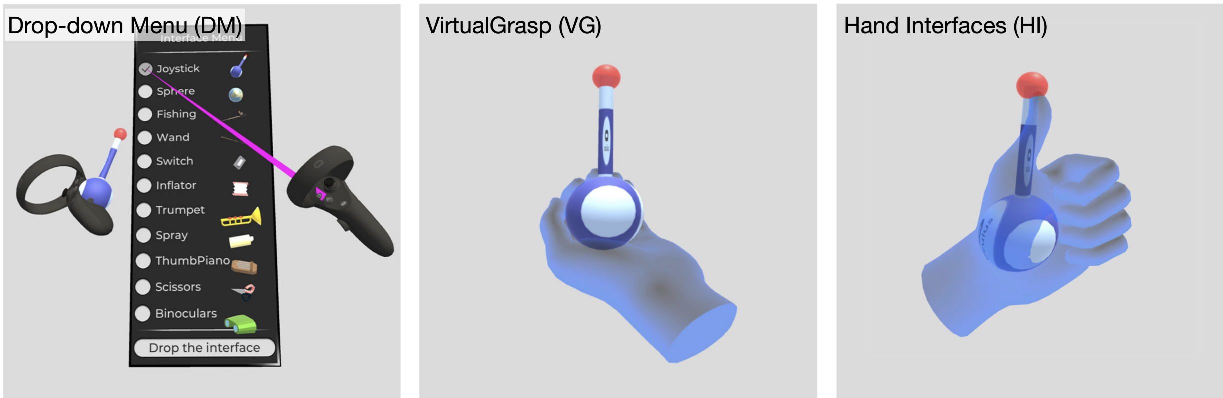

Figure 5.1: In this study we explored three techniques for the object retrieval scenarios, including two baselines and hand interfaces. The figure shows what retrieving a joystick looks like with each technique, from left to right: Drop-down Menu, VirtualGrasp, and Hand Interfaces.

by prior work, and were agreed on among researchers in this paper. 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 4.1) for VG and HI with different sets of gestures.

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. Upon conducting the study, the user's headset display was mirrored to a computer monitor to ensure the procedure was being followed properly.

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 a following section. Additionally, participants were asked about their rationales and any feedback they might have on each metric and design. Upon answering all questions, users moved on to the next object. Subsequent to retrieving all objects in a virtual environment, participants were loaded into the next environment to repeat the same process until all three were completed. The order of retrieval techniques and object retrieval were randomized. The study was completed in half an hour on average. As per COVID-19 regulations, masks were worn at all times throughout the study, participants were provided hand sanitizer, and the headset and controllers were cleaned between each user.

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 mean, standard deviation, p-value methods, and ran a thematic analysis with quotes from 17 participants. In this section, we report these findings, as well as additional insights that are not necessarily backed by the majority, but we still found to be interesting.

5.3.1 Quantitative Feedback

Participants were asked to provide scores on a 7-point Likert scale on how participants agreed that the design performed well by certain metric (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 provides feedback such that the virtual experience is similar to retrieving the object in reality. Freedom of movement refers to how unrestricted participants' movement is during retrieval. Swiftness of retrieval metric quantifies how quickly participants can retrieve an object from the standby hand pose (i.e., relaxed hand). Comfort of retrieval refers to how physically and mentally comfortable participants feel during retrieval. Finally, ease of recollection describes how easily participants can recall how to retrieve objects.

55 data points were collected per user (i.e., 5 data points $\times$ 11 interfaces) resulting in 935 data points in total across the 17 users. Figure 5.2 shows scores on all metrics averaged across participants, with error bars denoting standard deviations (i.e., variance across participants' feedback).

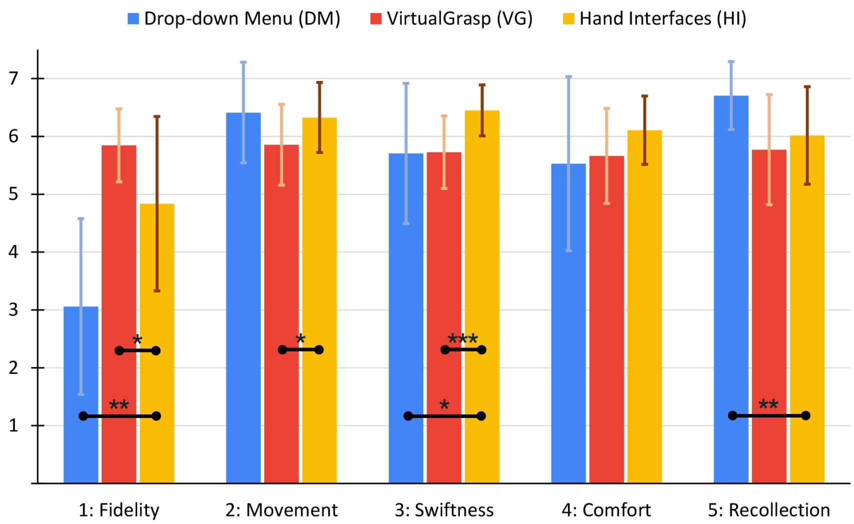

Figure 5.2: Retrieval Evaluation of Drop-down Menu, VirtualGrasp, and Hand Interfaces. The first (Blue), second (Red) and third (yellow) bar in each bar set indicates averaged scores of DM, VG, and HI techniques, with 1 being "strongly disagree" and 7 being "strongly agree" on a 7-point Likert scale. The number of asterisks denotes degree of significance between HI and baseline techniques. From the figure, we learn HI received significantly higher score than DM in fidelity during retrieval.

We conducted a significance analysis between HI and baseline techniques. Figure 5.2 visualizes different degrees of significance using asterisks ("$*$" denotes $p < 0.05$, "$**$" $p < 0.01$, and "$***$" $p < 0.001$). HI received higher score than DM in fidelity during retrieval. Meanwhile, participants perceived VG as a superior technique than HI in fidelity, which makes sense because VG leveraged people's experience of grasping objects in reality while HI chose not to. Another reason might be that, unlike interactive controls, object retrieval did not best leverage the merits of the tactile feedback in HI. For the movement metric, results indicated that HI was comparable with DM, and was considered better than VG. For the swiftness metric, participants preferred HI over the two baselines. The three interaction techniques were comparable in comfort, with no significant differences found. Finally, results indicated that it was least difficult to recall how to retrieve objects in DM, and HI was comparable with VG.

In addition to the overall comparison between HI and the two baseline techniques, we also looked into results for 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 5.3 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 color bars reflect participant percentages.

This visualization reveals the differences across designs – not only on score averages, but also on divergences. Additionally, participants' feedback on HI was more divergent than VG in general. This divergence was also reflected in participants' qualitative feedback, which we will discuss next. In HI, the three designs that received the most positive feedback – *Globe*, *Scissors*, *Binoculars* involved gestures people commonly make in reality, often used to imitate objects, which helped in participants' perception of realism in HI. 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 the participants reported.

For VG, participants' perception of realism of grasping certain objects depended on 1) whether participants grasped those objects in reality, 2) whether participants knew how to grasp those objects, and 3) whether participants grasped those objects the same way as in

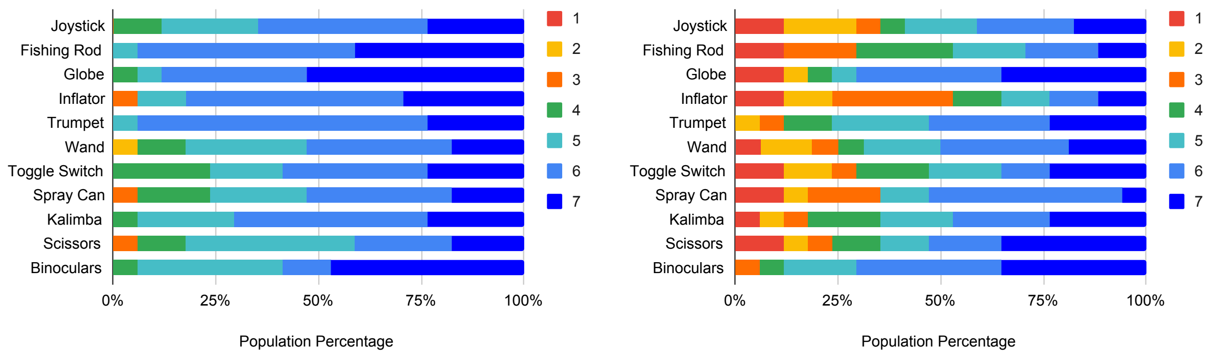

Figure 5.3: Score percentages of VirtualGrasp (the second baseline) and Hand Interfaces per Interface on Fidelity. Seven colors denote score on the mentioned 7-point Likert scale. The length of each color bar illustrates the number of users giving each score. The percentages are shown in horizontal axis while interface-wise user preferences are visible across each row. From the figure, we know that in HI, Globe, Scissors, Binoculars that involved gestures people commonly make in reality received the most positive feedback.

our study. Specifically, 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 did not know if the *Kalimba* is realistic. This lack of real-world experience also undermined the ease of recollection, as we found that 4 out of 17 participants had comments that linked their perception of realism with the ease of recollection. Furthermore, we saw a divergence of 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 in the perception of realism. P17 commented on the *Scissors* that "I feel the scissors is realistic because I am familiar with the rock paper scissors poses". 4 participants commented similarly. Other examples of this include the *Wand* and the *Binoculars*, the poses of which have already been commonly used in the real world. Compared with VG, participants did not focus on if the hand poses matched how they grasp objects in reality, due to knowing that HI was intended to allow them to imitate the object as opposed to grasp the object. 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, P1 commented that "I think kalimba is also realistic because the finger keys are kind of like a glove." Finally, P16 also commented that "If I can imagine my hand as an object, it is more realistic. I can imagine my hand as kalimba, but for other objects, it is more difficult". Finally, P13 commented "Sphere looks weird because it is like my hand is wrapped with the globe. However, if I think it of my hand as becoming a globe, it seems very realistic. It depends on how I think about it."

5.3.3 Qualitative Feedback — Visuals on the Perception of Realism

We found that visual feedback also had a major impact 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 impact. Interestingly, we observed something similar to the uncanny valley effect — we found that small unrealistic elements (i.e., discrepancies between users' mental model and what they saw) could severely impact the perception of realism. This effect, in some cases, makes VG less favourable compared with HI because participants often had 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 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* example 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 with their experience with the physical world. One example of this, quoted from P12, said that "Hand interfaces is not realistic because it feels odd for objects to go through my hand". Some participants suggested hiding 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 cases 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 completed.

5.3.4 Qualitative Feedback — Ease of Recollection

We found that the lack of real-world experience, which resulted in confusion on realism, could undermine the ease of recollection as we noted earlier that 4 out of 17 participants had comments that linked ease of recollection with perception of realism. In other words, if participants had never used or seen usages of 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 — the ambiguity in cases where there were multiple different hand poses to grasp an object, and where 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 *Fishing rod*, making both difficult to recall. Similarly, P16 found that the hand pose to retrieve the switch was similar to how people grasp many small objects, therefore making it difficult to recall. In addition, 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 on hand poses being ambiguous. We are cautious that this might be due to our selection of designs in the study. Nonetheless, we believe that the ambiguity problem of HI is much less than VG for the fact that there are simply more objects of various shapes that can be grasped with similar hand poses (i.e., ones that are challenging to disambiguate in VG) than objects of similar shapes that can be grasped with different hand poses (i.e., ones that are challenging to disambiguate in HI). However, it was obviously not a slam dunk for HI. Though the ambiguity was minor, one participant expressed concerns that there might simply be too many things in the real world, which will be challenging to remember if we want to make them into interfaces (P4). Overall, our observation is that having too many things to remember is just as difficult as having very similar things to remember. However, each technique has unique pros and cons that compensate for one another, and it would be beneficial to consider both in AR/VR interaction designs.

5.3.5 Qualitative Feedback — Comfort

Finally, we have seen a discussion point around comfort. Four participants found several hand poses in VG uncomfortable. Among these participants, three found the angle of specific fingers (i.e., pinky in *Trumpet* and *Spray can*) or the wrist (i.e., *Kalimba*) awkward to perform. Two participants mentioned that the *Spray can*, *Joystick*, and *Inflator* could not be maintained for a long time due to fatigue. Nonetheless, we did not notice a significant difference between participants' quantitative feedback on the comfort of the two designs. What we inferred from this result was that while HI designs were comfortable to most users, comfort level was 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.

CHAPTER 6

Study of Interactive Control

With the same set of participants and virtual objects, we conducted a second study investigating users' observation 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 6.1 left). We selected this gesture because we found the fist to be one of the simplest gestures for people to perform. This gesture simply served as an anchor point without considering objects' affordances or shapes. 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., [9]), as well as prior work (e.g., [43]) where users interact with virtual objects by directly manipulating them with their hands (Figure 6.1 middle). This direct manipulation technique includes moving parts of objects (e.g., *Joystick*), touch sensing (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).

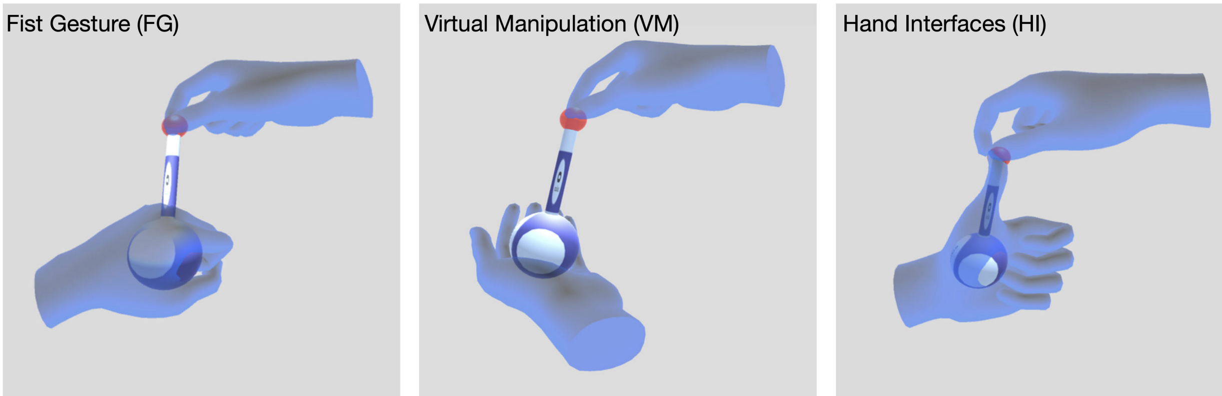

Figure 6.1: In this study we explored three techniques for the interactive control scenarios, including two baselines and hand interfaces. The figure shows what interacting with a joystick looks like with each technique, from left to right: Fist Gesture, Virtual Manipulation, and Hand Interfaces.

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 disabled object retrieval in the software. Specifically, the study started with virtual objects already coupled with users' hands and experimenters switched between designs by pressing shortcut keys on a physical keyboard.

6.2.1 Procedures

The procedure for interactive control was mostly the same as that of the 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, which we designed to have simple interactive mechanisms (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, same as the ones in the previous study. This study took about half an hour to finish.

6.3 Evaluation Results

6.3.1 Quantitative Feedback

We used the same set of metrics in this study as the ones used in object retrieval study but their meanings were slightly different as the interaction scenarios had changed. Fidelity of interaction refers to 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 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 interaction. The last metric, ease of recollection, describes how easily participants can recall how to do interactive control with objects.

In total, we collected the same amount of data points as the object retrieval study. Figure 6.2 shows the result.

Referring to Figure 6.2, HI received a higher average score in interaction fidelity. Participants commented that the lack of tactile feedback in FG had a negative impact on their scores. Interestingly, while VM also lacked tactile feedback, it had a higher average score than HI (no significance). This was possibly due to participants considering grasping objects as part of the interaction as it was challenging to not consider both hands in bi-manual interactions holistically. The results indicated that having a realistic grasping gesture contributed similarly as having tactile feedback.

For the movement metric, we found no statistical differences between FG and HI or between VM and HI. However, HI had the lowest average score of the three. This is likely due to some HI interactions being naturally restricted by the hand's degree of motion whereas

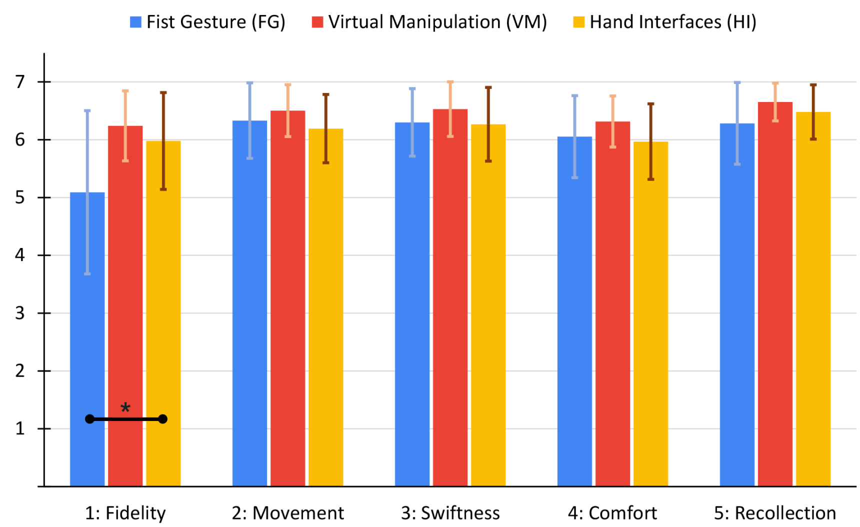

Figure 6.2: Interaction Evaluation of Fist Gesture, Virtual Manipulation, and Hand Interfaces. The first (Blue), second (Red) and third (yellow) bar in each bar set indicates averaged scores of FG, VM, and HI techniques, with 1 being "strongly disagree" and 7 being "strongly agree" on a 7-point Likert scale. The number of asterisks denotes degree of significance between HI and baseline techniques. From the figure, we learn HI received higher score than FM in fidelity during interactive control with relative significance.

FG and VM interactions are unrestricted. For example, both the FG and VM interaction for the *Fishing rod* involved simply grasping the virtual reel in the air, whereas the HI interaction required pinching the thumb, thus resulting in the latter interaction being restricted by the thumb's degree of motion.

Additionally, we found no significance for the swiftness metric, comfort, and recollection. Nonetheless, HI received the lowest score among the three and some participants found moving joints and fingers to be uncomfortable in long-term use, for which we are cautious and hope to conduct future research to investigate. Some participants noted that the tactile feedback in HI helped in the recollection of interactions. We will discuss this feedback in the qualitative results section in detail.

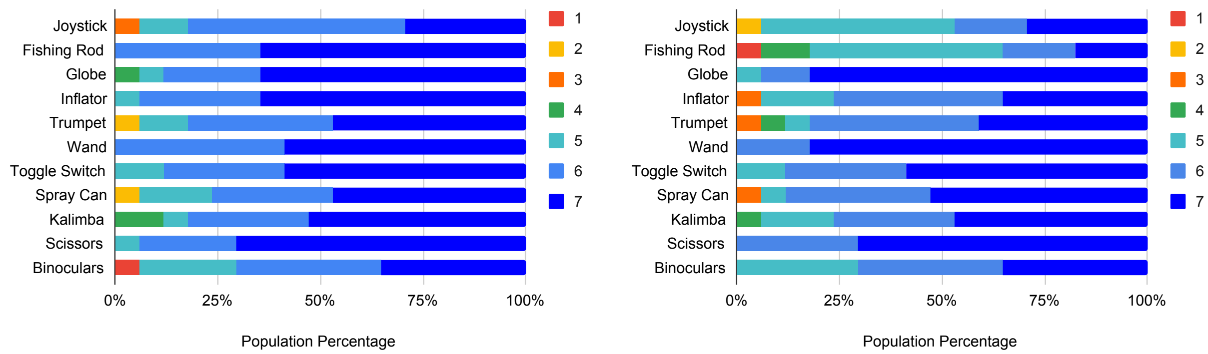

Figure 6.3: Score percentages of Fist Gesture (the first baseline), and Hand Interfaces per Interface on Freedom of Movement. The figure indicated a clear preference for the FG Fishing rod over that of the HI with respect to freedom of movement.

We also examined the design-wise score percentages for all metrics and all interaction techniques, which can be found in the appendix. Figure 6.3 shows two examples for the movement metric for FG and HI. Figure 6.3 left, shows that the FG *Fishing rod* received a majority score of 7 by 11 participants while Figure 6.3 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 the HI with respect to freedom of movement. FG interaction, as we specifically designed, was least restricted by the anatomical restraints of the hand among all interaction techniques. In contrast, the HI interaction required pinching the thumb and therefore constrained the interaction to the degrees of freedom the thumb joint offered, ultimately reflecting in the lower movement score. A similar case was exhibited with the *Joystick* scores but they were less affected compared with the *Fishing rod* due to the fact that the *Joystick* is more kinematically similar to the thumb than a rotating reel.

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

Because interactivity is an additional step after object retrieval, participants' perception of interactions inevitably included their prior perceptions of object retrieval or at least the perception of the coupling between virtual objects and their hands (a major part of object retrieval). The knowledge generated in the previous section largely applies to this section. Here, we focus on new knowledge generated from interactions (i.e., use of objects) as opposed to from object retrieval.

One main factor which we have seen on participants' perception of realism was around uni-manual vs. bimanual 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 virtual reality had more impact on their perception of realism than tactile feedback. In this regard, we learned that, if not designed in a way that is reasonably consistent with participants' real-world experience (i.e., *Spray can*), HI designs could be perceived as unrealistic despite the tactile feedback.

Another insight we drew from one participant's comment is that the tactile feedback on the left hand could break the immersion participants previously reported helped with their perception of realism. In other words, the more participants are aware of their left hands, the more difficult it becomes to imagine their hands becoming objects.

Overall we found tactile feedback having positive impact on participants' perception of realism. One participant commented on the *Joystick* that having something tangible to grasp made it easier for the manipulating hand to maintain its position, which was more like what they do in the real world. One participant praised the tactile feedback in HI for making the design feel more realistic. The same participant, as well as eight other 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". P11 mentioned that "For tasks requiring much interaction, I prefer HI because of its tangible feedback". Finally, the average score for fidelity comparatively improved by a large margin in interactive control scenarios from in object retrieval scenarios, for which we suspected that the tactile feedback played a large role.

6.3.3 Qualitative Feedback — Tactile Feedback

Tactile feedback also received much commentary from participants other than implications on the perception of realism. Six participants found various reasons that tactile feedback benefited their interaction with virtual objects. P10 preferred the *Kalimba* design in HI over VM, for the reason that they could better detect key presses since their fingers (both manipulating and imitating) were able to feel the tactile feedback. P3 felt that the control of the *Joystick* was better with something tangible so that they could grasp something as opposed to nothing in the air. 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 commented that "keeping hands stable makes me feel comfortable". Additionally, we have also 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, it became obvious for participants to know how to interact with the objects as participants have used many of these objects in the real world. In the study, we found participants relied solely on visuals 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 about the couplings between these objects and users' hands. However, one interesting point we noted was that participants found interacting with an object later helped them recall the object retrieval more effectively. One participant mentioned that the *Inflator* hand pose suddenly made a lot of sense and therefore was easy 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 from hand kinematics. At the end of the day, our knuckles and joints can only move so much before we experience discomfort. Our hands are constrained but the objects they are supposed to imitate might not be. 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 a very interesting design opportunity of using visual illusion or scaling (e.g., [10, 21]) to compensate for limited range of motion in HI. For example, a 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 to play in several cases that participants found favorable. For example, HI allowed participants to use their index finger as a wand, which one participant found particularly useful for being able to point at things more precisely than techniques that asked the participant to grasp objects (nothing but air in reality). Latter techniques required users to rely on mostly visuals in pointing whereas HI leveraged proprioception of users' fingers.

Social acceptance. Social acceptance was another discussion point. One participant commented that the hand pose with the pinky finger up has social implications and therefore they would not use it in public. Another participant suggested using the middle finger as a joystick, which we felt might be socially unacceptable in most cases. This divergence in these two examples reminds us that social acceptance might be more complicated than we thought, and it would be beneficial to consider both application contexts and user backgrounds when we optimizing for social acceptance in HI.

Fun to use. Even though we did not explicitly ask participants, two mentioned 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 user 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 VM. Not only did this ambiguity demand more effort to memorize, it also created challenges in detection. In fact, to make sure the technical performance did not interfere with the user study, we segmented the designs of VM into batches so that objects with similar retrieval gestures never showed up in the same batch. This effort successfully avoided false detection results without compromising the study flow since the transitions between batches were as quick as the experimenter pressing a key. However, we could imagine the ambiguity in detection demands additional input to resolve for developing object retrieval techniques with freehand grasping. On the other hand (or the same hand), HI often yields a more diverse set of hand poses that makes detection easier and thus interaction simpler.

Design space / Customizability. 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.

CHAPTER 7

Example Use Cases

7.1 Education

In educational scenarios (Figure 7.1), students can use Hand Interfaces to quickly and easily retrieve tools that facilitate their learning experiences. In this circuit example, a student raises up two index fingers to retrieve *Multimeter probes* to measure voltage of an important circuit component (A). After the measurement, the student uses the *Pen* to write down a note (B). Finally, the student takes a photo of the entire set up using the *Camera* (C).

7.2 Entertainment

Hand Interfaces can also be easily applied in entertainment applications. In a magic fighting game (Figure 7.2), players retrieve wands once the *Wand* gesture is performed (A). By waving their wands (index fingers) following specific trajectories, players can cast different spells as tactics to win the fight. After taking some hits, a player can retrieve a healing potion by performing a *Mug* gesture and use the potion by "drinking" from the mug (B). Players can also retrieve other tools, for example, a *Book*, and level up by opening the book (C).

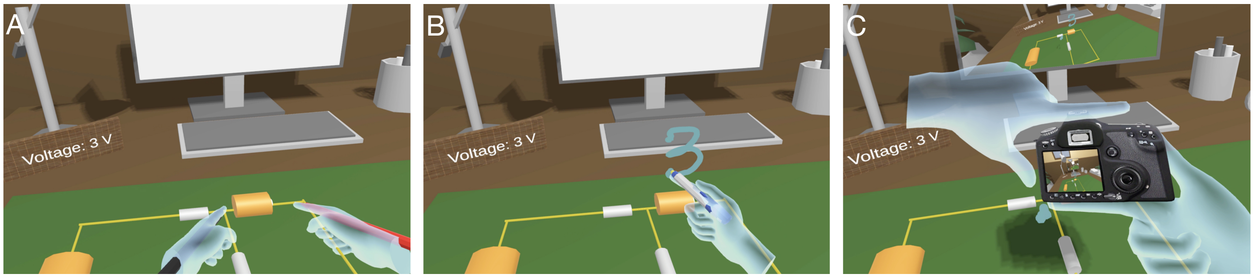

Figure 7.1: Hand Interfaces allows users to easily use various tools in an educational setting. Specifically, *Multi-meter probes* are used to measure voltages (A), the *Pen* is used for writing notes (B), and the *Camera* (C) is used for taking a photo of the current scene.

Figure 7.2: In a magic fighting game, a user uses the hand-imitated *Wand* to cast spells (A), the *Mug* to use healing potion (B), and the *Book* to level up (C).

7.3 Ubiquitous Computing

Finally, Hand Interfaces is directly applicable to AR scenarios, in which many applications demand hand-free interactions so that users can quickly switch between their tasks in the physical world and in the digital world (Figure 7.3). 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).

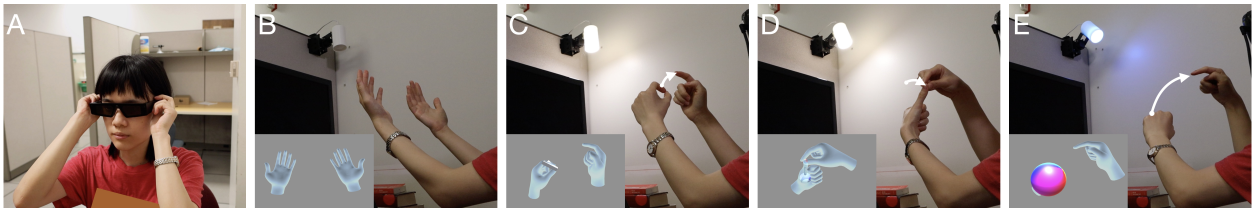

Figure 7.3: A user uses Hand Interfaces in concert with an AR device (A, B) to quickly and easily control a smart lighting system. Specifically, the *Toggle switch* is used to turn on/off the light (C), the *Joystick* controls its pan/tilt (D), and a fist-imitated *Color palette* is used to set its color (E).

CHAPTER 8

Discussion and Conclusion

First, Hand Interfaces does not support virtual objects that are too large or small compared with human hands. Future effort will look into visual illusion and scale (e.g., [10, 21]) 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. By addressing objects of different scales, the hand-based interaction framework in AR/VR will be more complete.

Additionally, we learned from the study that users often have strong needs on adjusting the system to their preferences. For example, Hand Interfaces in the future should allow users to easily define their own gestures, or offer them several options to choose from for imitating objects. It is technically achievable by designing a DIY-gesture UI for adding their customized gestures to existing gesture library. This direction of work will dive deeper into user-in-the-loop interaction design and accessibility, which builds upon but goes beyond current system, making the work more influential.

Our experiments suggest that a robust detection system requires fine-tuned sensitivity matrices (i.e. having lower weights for joints that move around during interactions), which were heuristically generated based on researchers' practices and piloting results. Future researcher should keep this in mind when using template matching for dynamic hand gesture detection. Fine-tuned sensitivity matrices are helpful for robustness, however, this approach might not be intuitive to potential designers with less programming experience. To resolve this issue, future researchers can envision an automated calibration process to generate sensitivity matrices, given the known kinematic characteristics of human hands and the imitated objects as input. Developing a toolkit with user interfaces that allow users to easily calculate and adjust parameters in a calibration process should be a system/framework research with enough novelty and significance.

There are technical challenges from the computer vision approach that powers Hand Interfaces' detection. For example, occlusion is unavoidable when two hands interact with each other and causes errors in hand tracking. To compensate for this system error, participants were asked to perform gestures that avoid occlusions in our studies. In addition, touch detection is innately challenging to cameras since differences between “touch” and ”no-touch” oftentimes are minute. Since our contribution falls in interaction techniques and evaluation, we did not push the performance to the limit. Specialists in computer vision can move forwards for superior technical approaches addressing these challenges, as some existing research projects have already shown promise [38, 58]. If the problems are solved, the prototypes in this paper will turn into mature commercial products as new widely-used interface for AR/VR, just as how useful a mouse can be for a 2D computer screen.

From the perspective of interaction design, Hand Interfaces introduces a new way to interact with the virtual world as well as an inspiring thread of thinking for future researchers. Existing interaction techniques focus on leveraging design ideas that people have experience of. For example, the idea behind Drop-down Menu (i.e., the baseline technique in our study) is the same as the GUIs on conventional computer platforms. It leverages user experience of using these digital interfaces. Meanwhile, direct Virtual Manipulation (e.g., [56, 9]) opens up a new direction, which attempts to reproduce user experience of manipulating physical objects in reality (i.e., physical interfaces). Extending user experience from conventional interfaces (i.e., digital and physical) to 3D virtual worlds utilizes prior knowledge that improves learnability, but it might miss the large design space AR/VR uniquely provides — Hand Interfaces do not primarily attempt to leverage users' real-world experience, but instead creates a new interaction modality specific to AR/VR by allowing users to have their hands become these objects through imitation.

In 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 is effective, expressive, and fun to use. Finally, we have shown example applications centering around three domains — entertainment, education, and ubiquitous computing. All these efforts have been open sourced to facilitate future research.

APPENDIX A

Design-wise Score percentages Diagrams

A.1 Retrieval Evaluation of Drop-down Menu (the First Baseline)

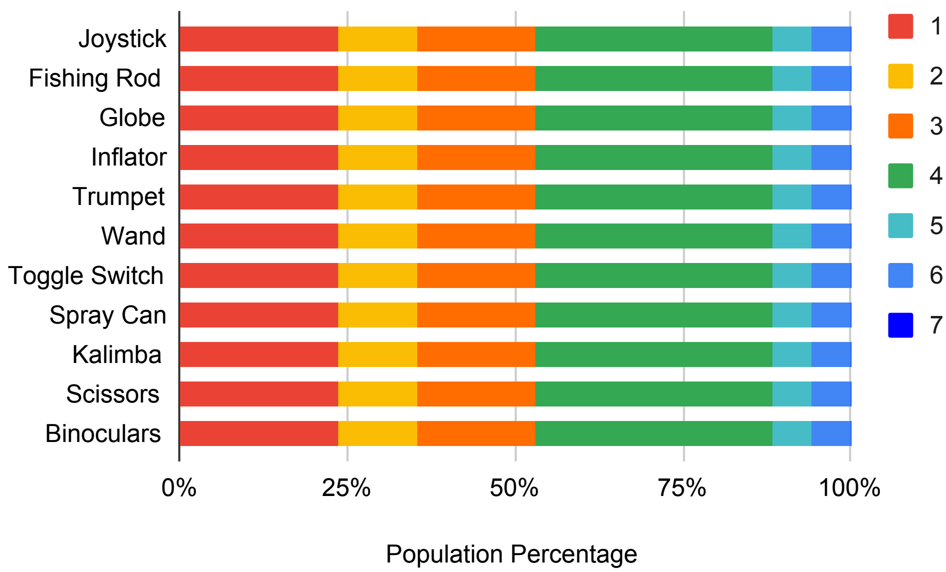

Figure A.1: Retrieval Evaluation of Drop-down Menu per Interface on Fidelity

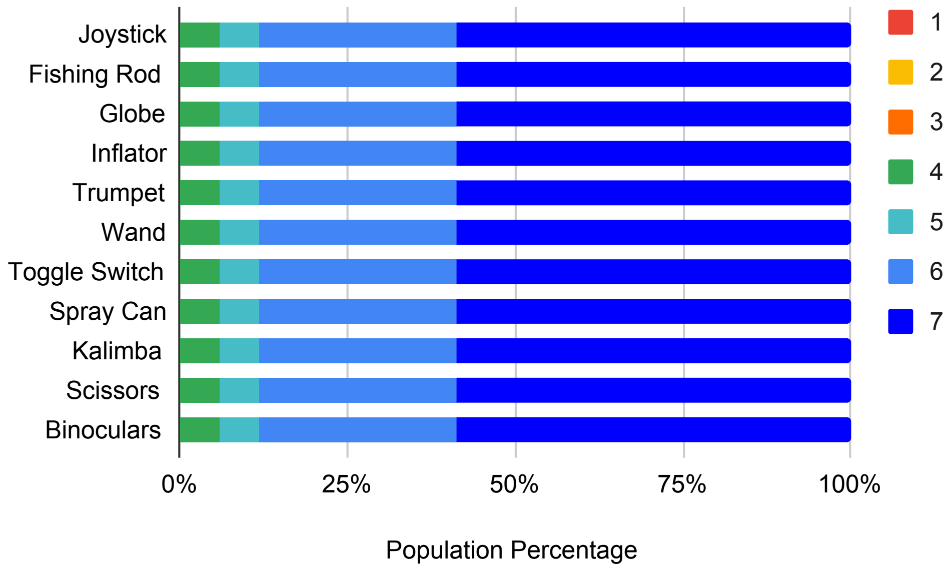

Figure A.2: Retrieval Evaluation of Drop-down Menu per Interface on Movement

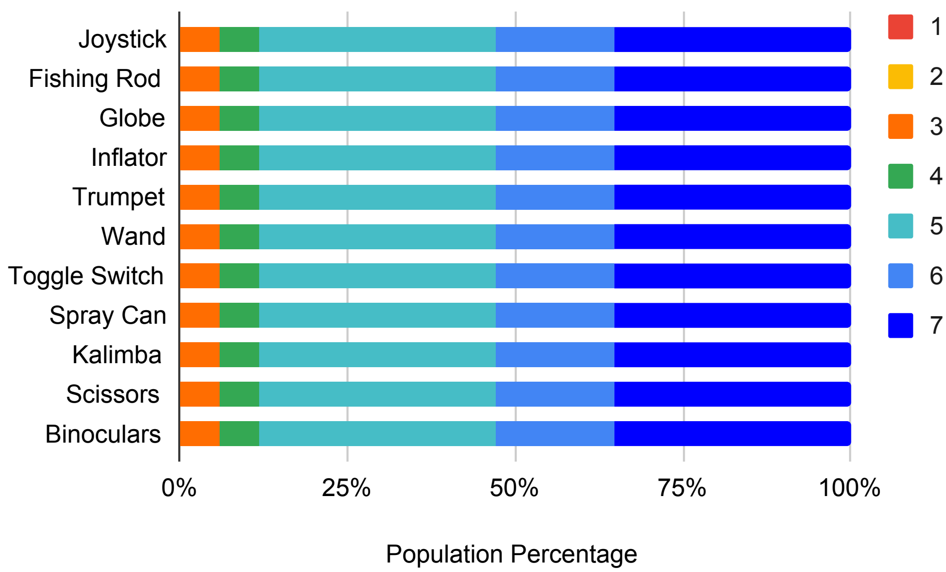

Figure A.3: Retrieval Evaluation of Drop-down Menu per Interface on Swiftness

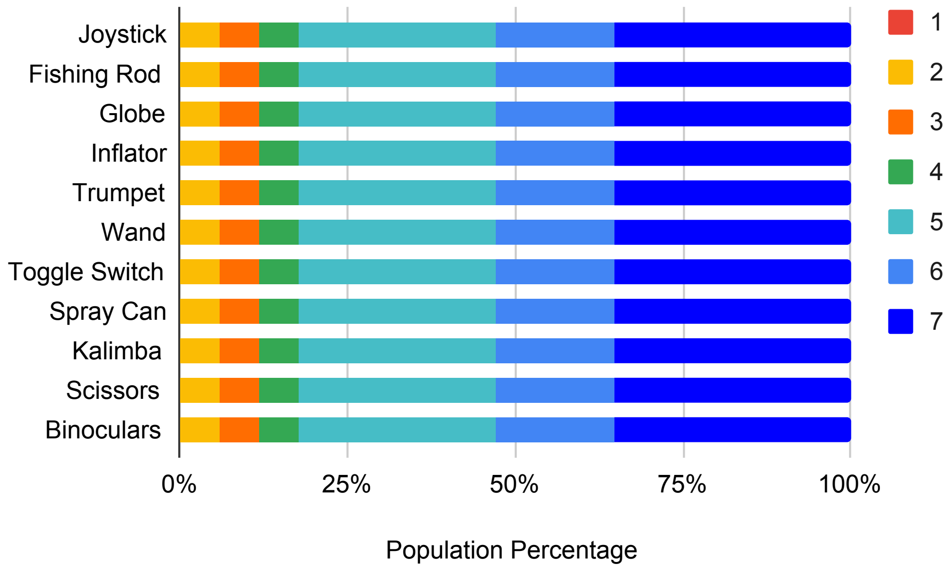

Figure A.4: Retrieval Evaluation of Drop-down Menu per Interface on Comfort

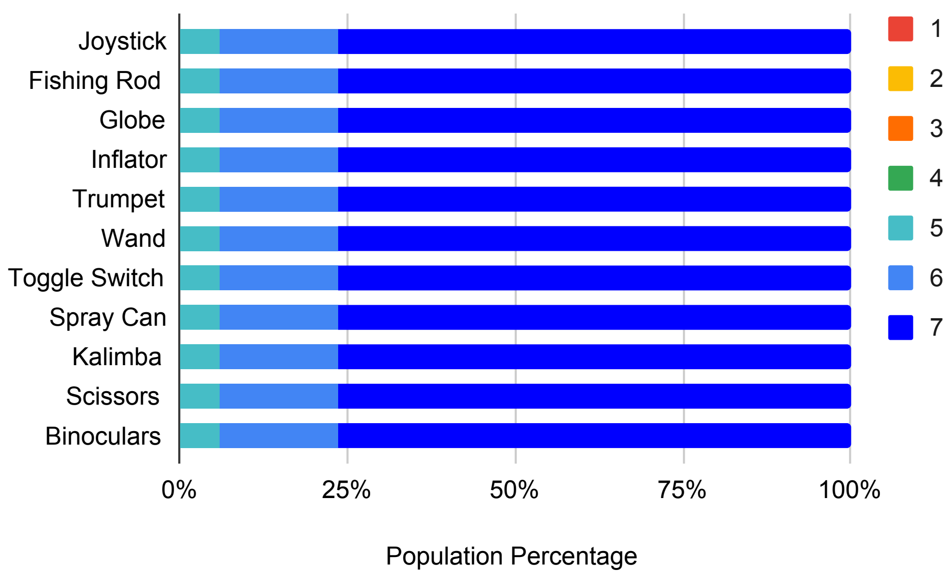

Figure A.5: Retrieval Evaluation of Drop-down Menu per Interface on Recollection

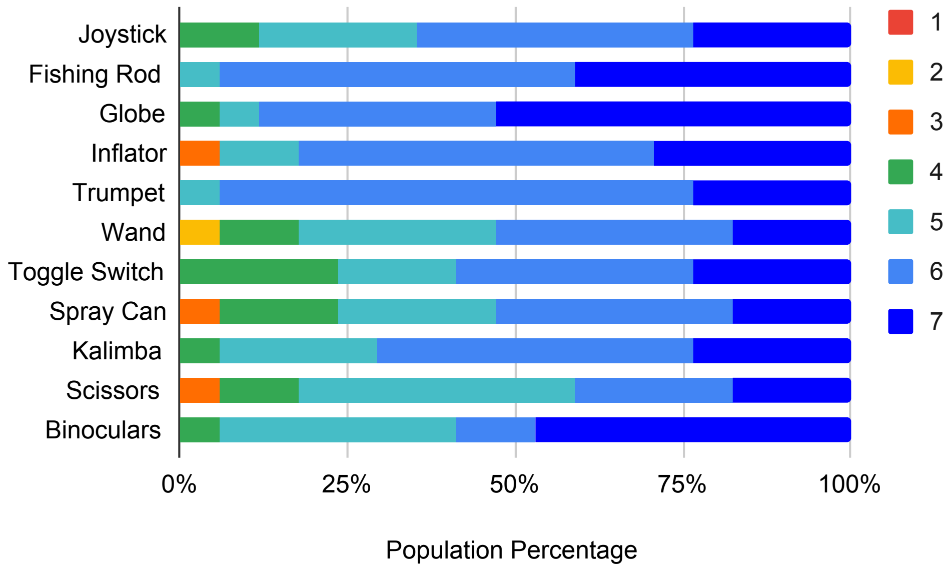

Figure A.6: Retrieval Evaluation of VirtualGrasp per Interface on Fidelity

A.2 Retrieval Evaluation of VirtualGrasp (the Second Baseline)

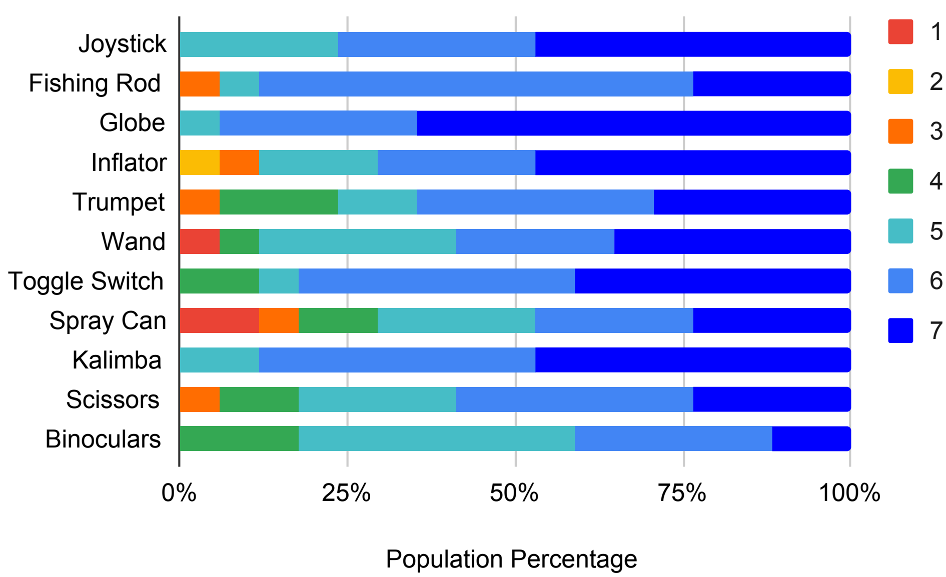

Figure A.7: Retrieval Evaluation of VirtualGrasp per Interface on Movement

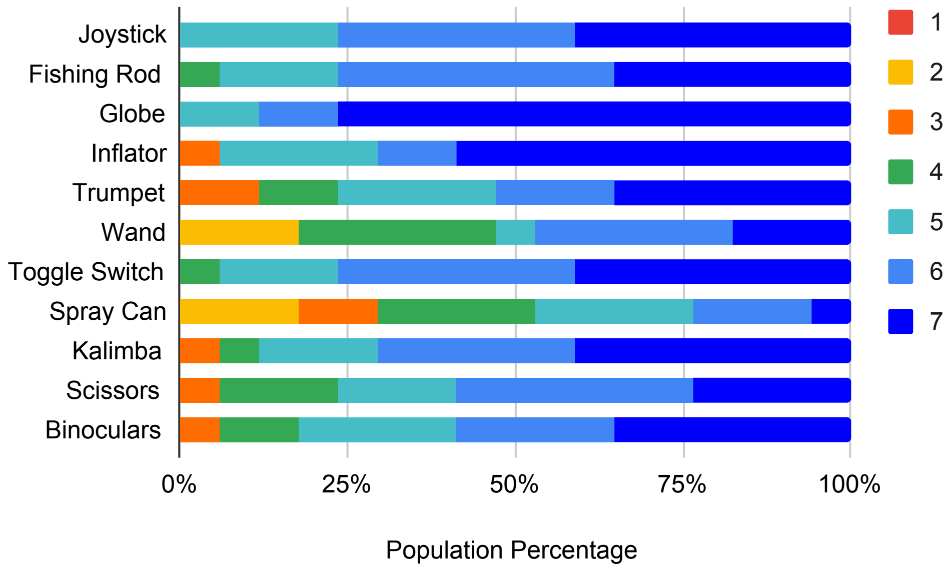

Figure A.8: Retrieval Evaluation of VirtualGrasp per Interface on Swiftness

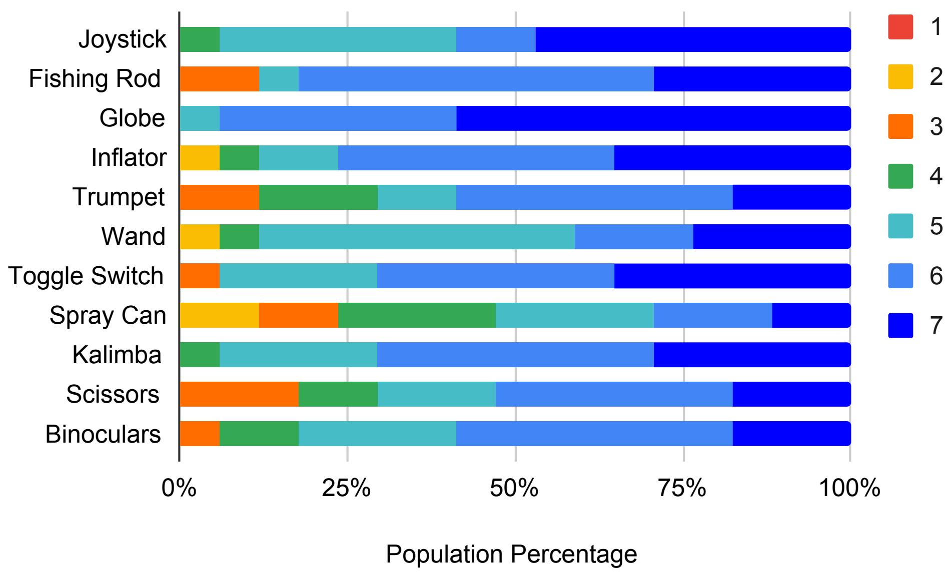

Figure A.9: Retrieval Evaluation of VirtualGrasp per Interface on Comfort

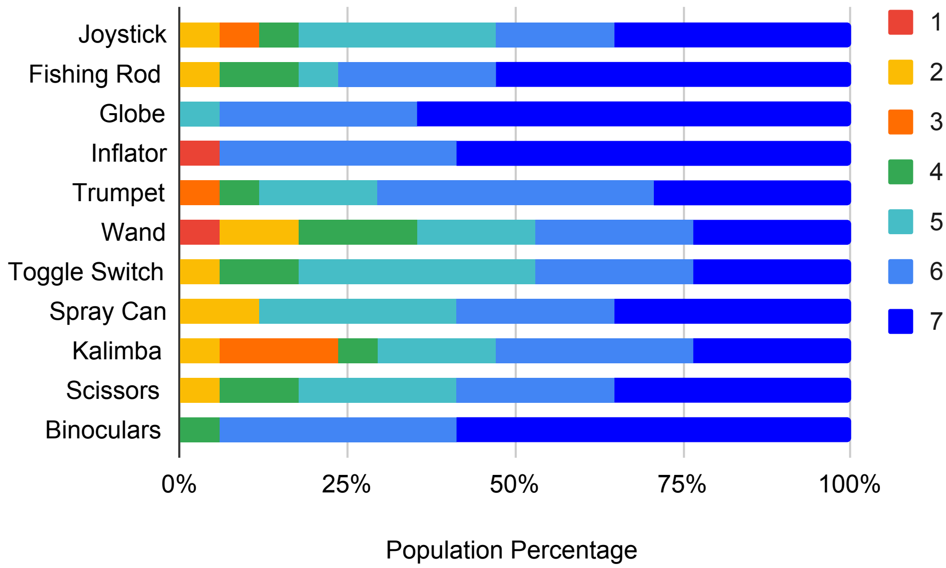

Figure A.10: Retrieval Evaluation of VirtualGrasp per Interface on Recollection

A.3 Retrieval Evaluation of Hand Interfaces

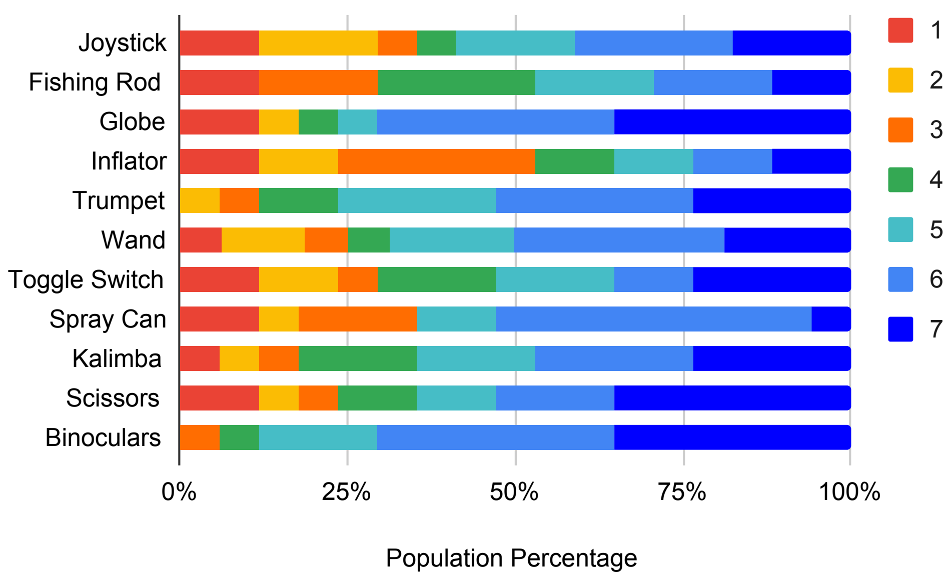

Figure A.11: Retrieval Evaluation of Hand Interfaces per Interface on Fidelity

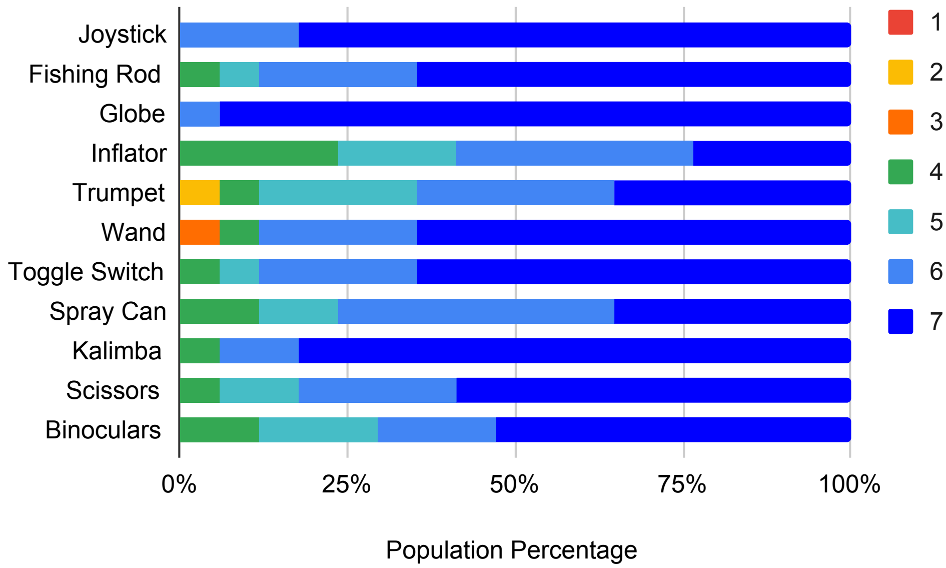

Figure A.12: Retrieval Evaluation of Hand Interfaces per Interface on Movement

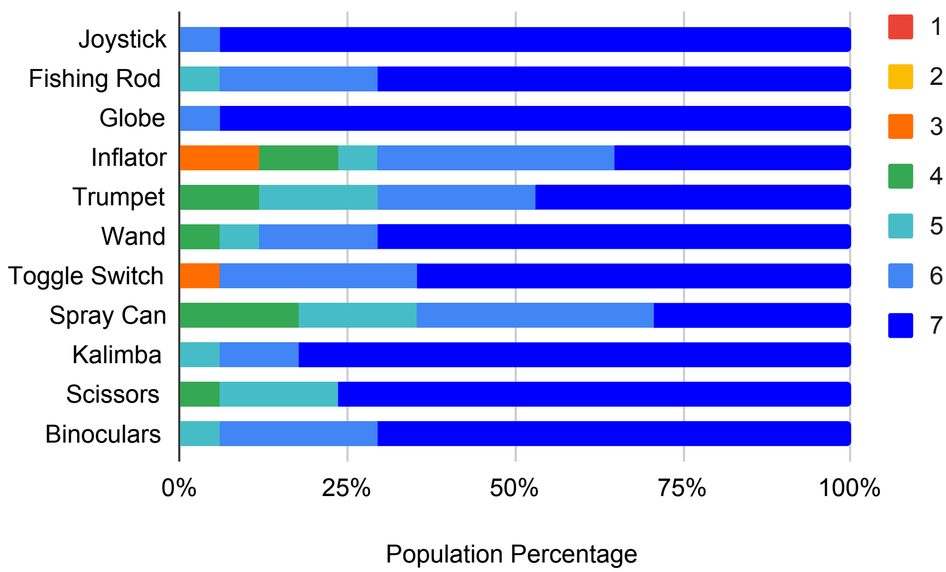

Figure A.13: Retrieval Evaluation of Hand Interfaces per Interface on Swiftness

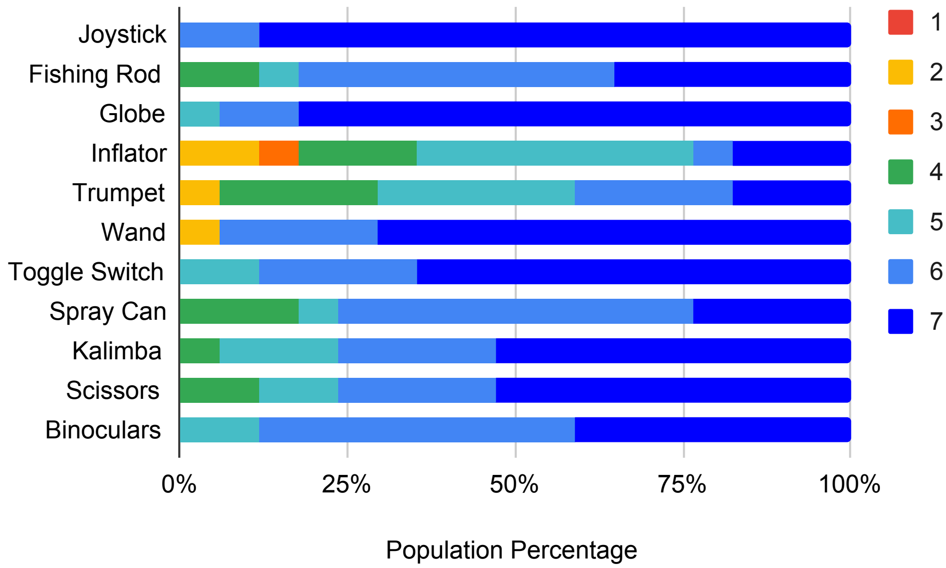

Figure A.14: Retrieval Evaluation of Hand Interfaces per Interface on Comfort

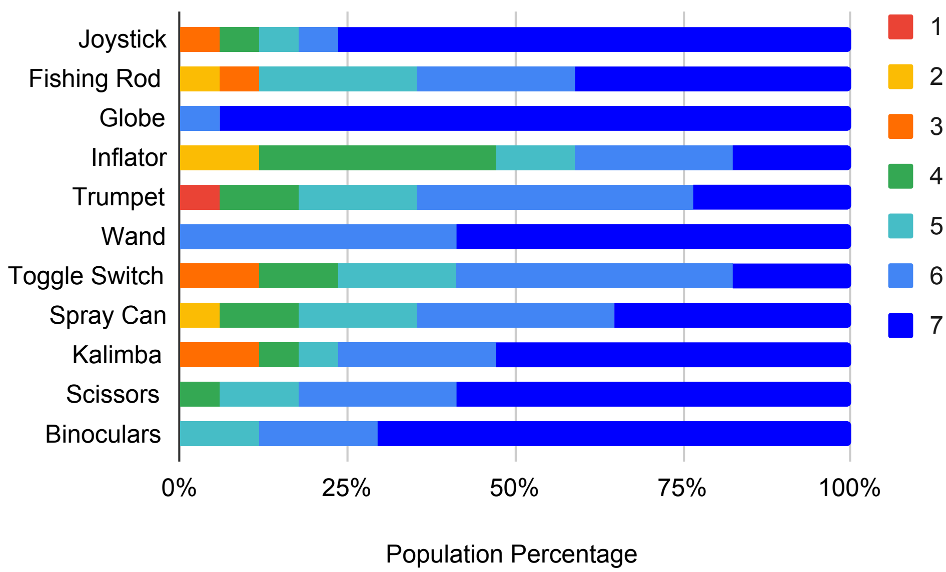

Figure A.15: Retrieval Evaluation of Hand Interfaces per Interface on Recollection

A.4 Interactivity Evaluation of Fist Gesture (the First Baseline)

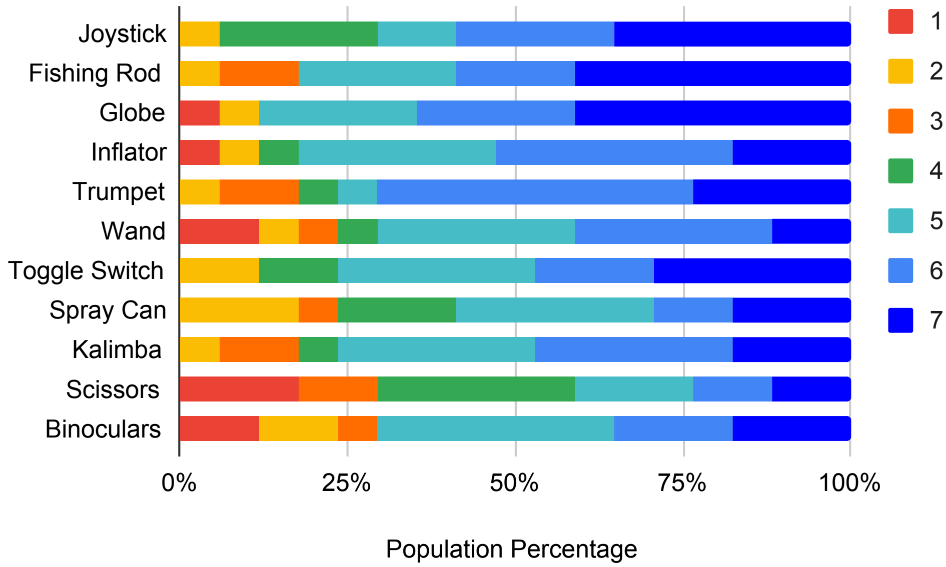

Figure A.16: Interactivity Evaluation of Fist Gesture per Interface on Fidelity

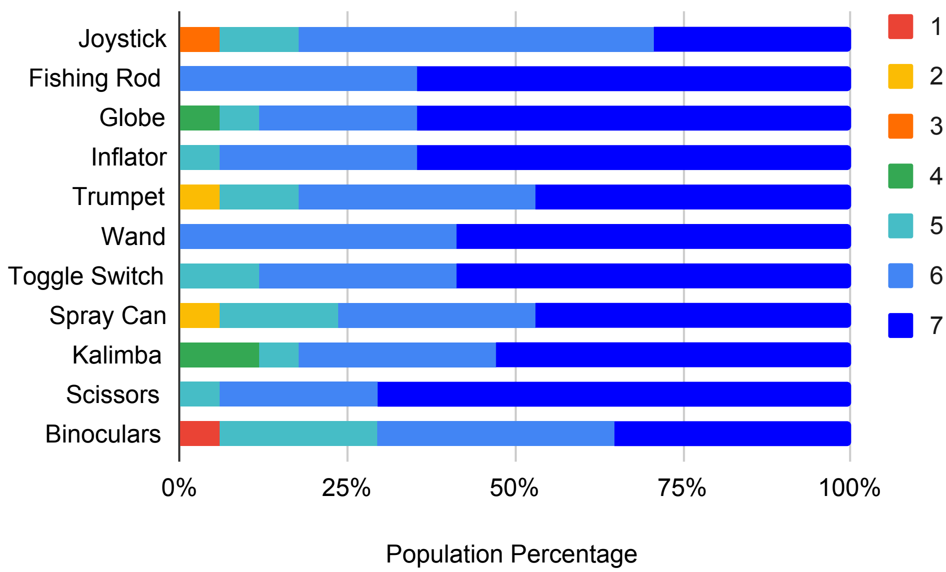

Figure A.17: Interactivity Evaluation of Fist Gesture per Interface on Movement

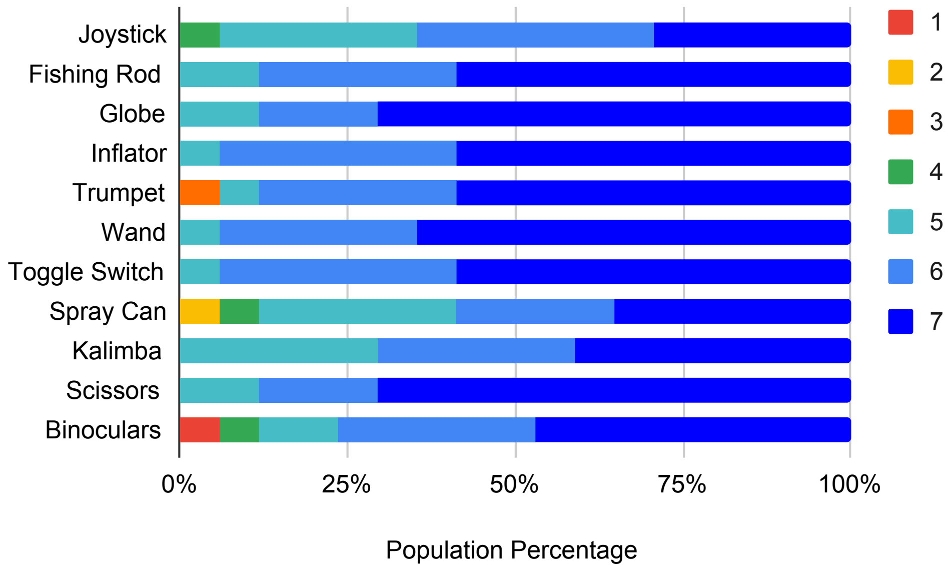

Figure A.18: Interactivity Evaluation of Fist Gesture per Interface on Swiftness

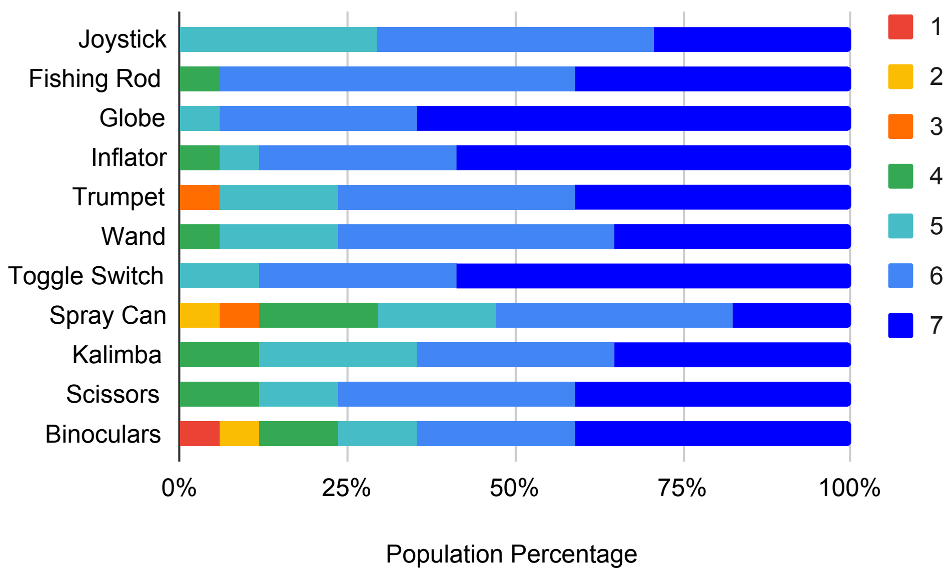

Figure A.19: Interactivity Evaluation of Fist Gesture per Interface on Comfort

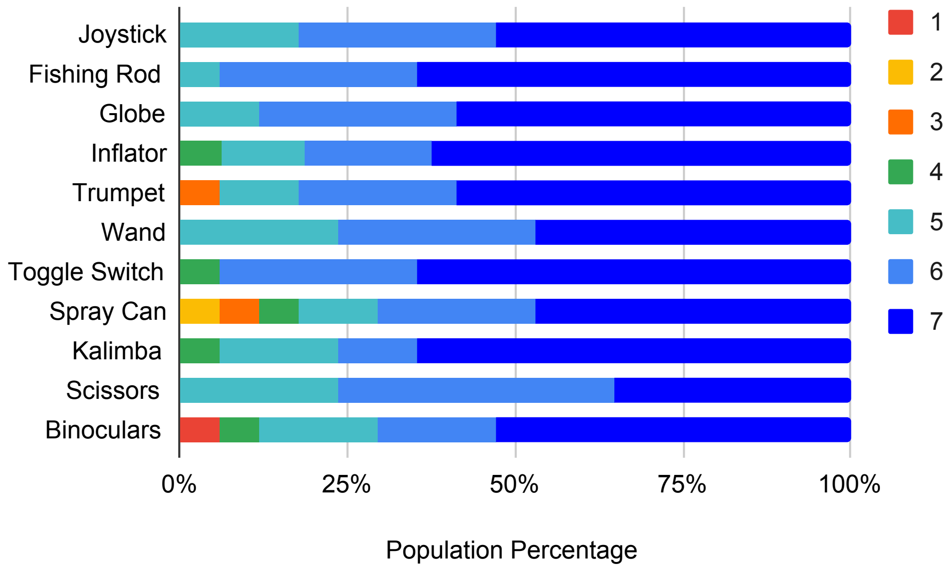

Figure A.20: Interactivity Evaluation of Fist Gesture per Interface on Recollection

A.5 Interactivity Evaluation of Virtual Manipulation (the Second Baseline)

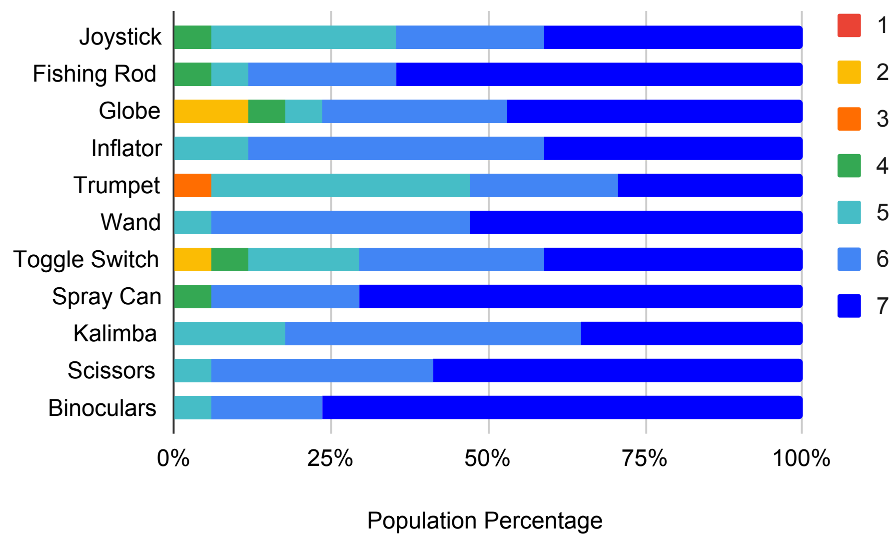

Figure A.21: Interactivity Evaluation of Virtual Manipulation per Interface on Fidelity

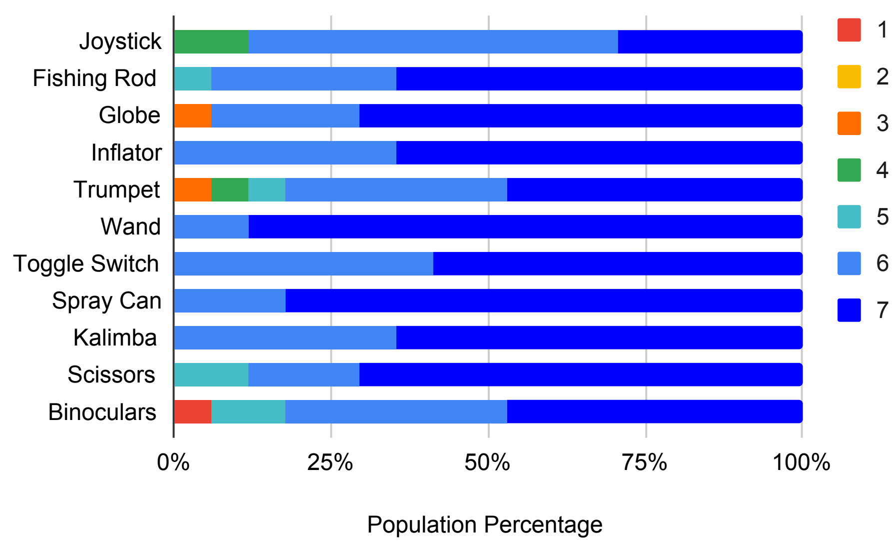

Figure A.22: Interactivity Evaluation of Virtual Manipulation per Interface on Movement

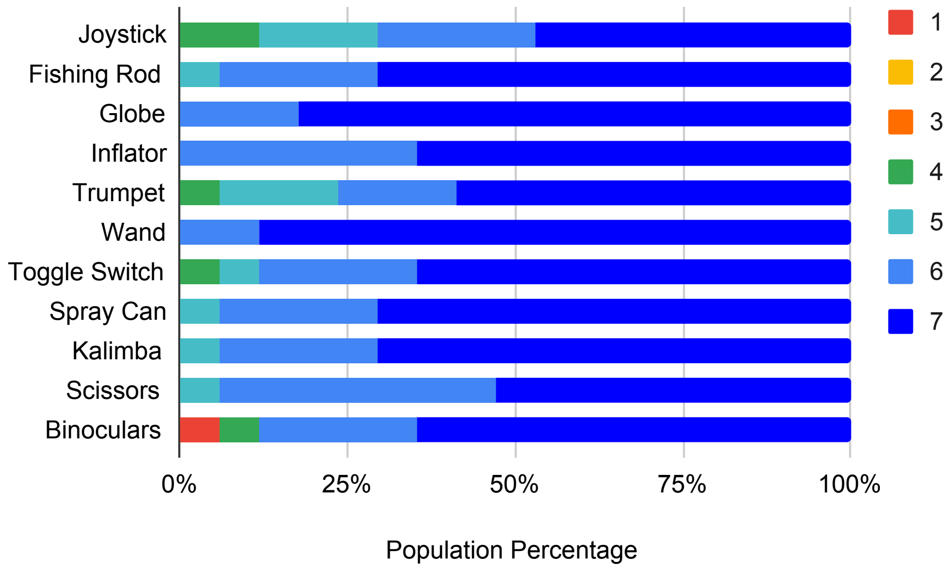

Figure A.23: Interactivity Evaluation of Virtual Manipulation per Interface on Swiftness

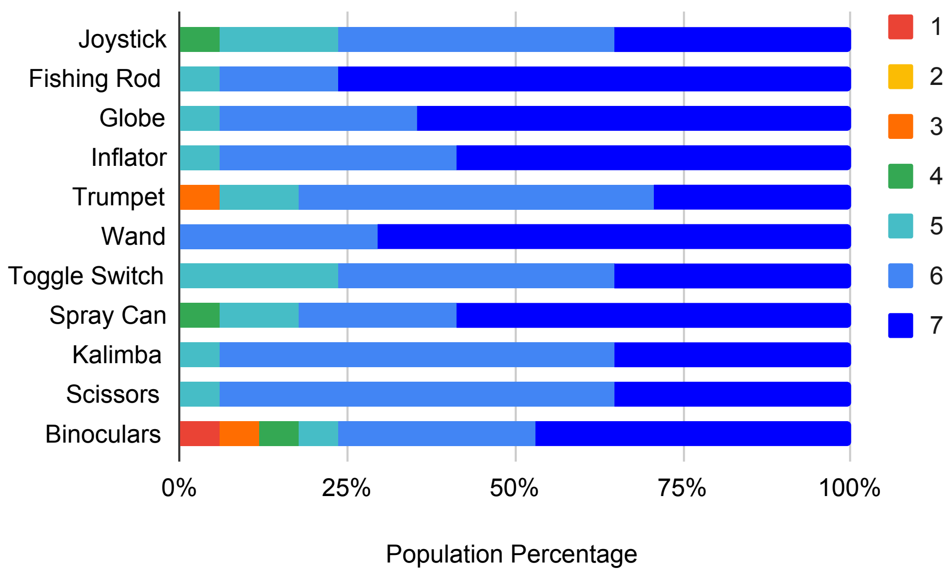

Figure A.24: Interactivity Evaluation of Virtual Manipulation per Interface on Comfort

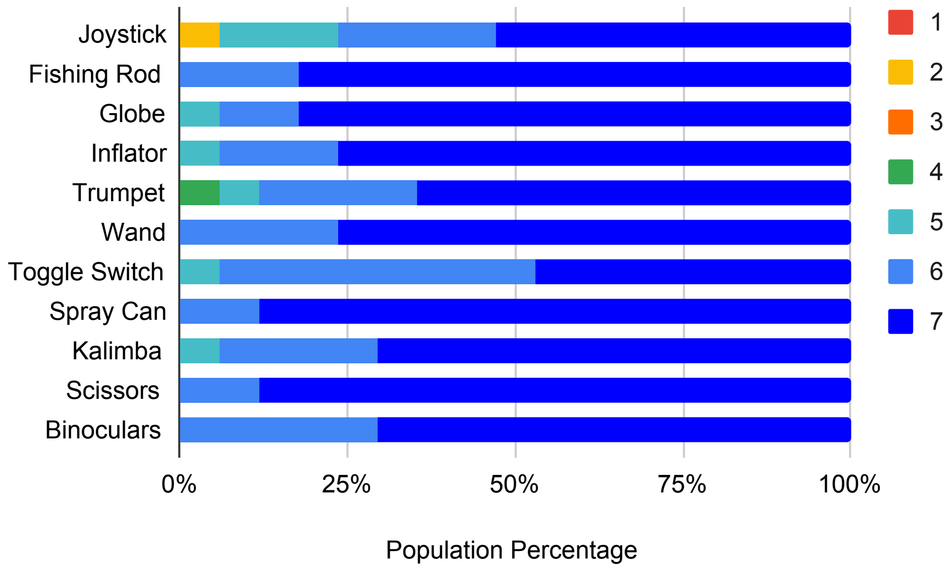

Figure A.25: Interactivity Evaluation of Virtual Manipulation per Interface on Recollection

A.6 Interactivity Evaluation of Hand Interfaces

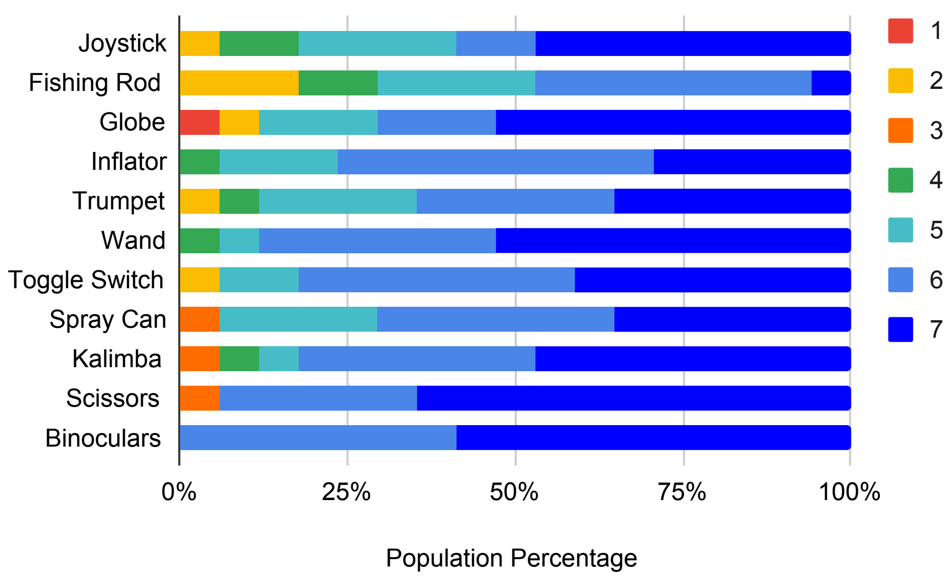

Figure A.26: Interactivity Evaluation of Hand Interfaces per Interface on Fidelity

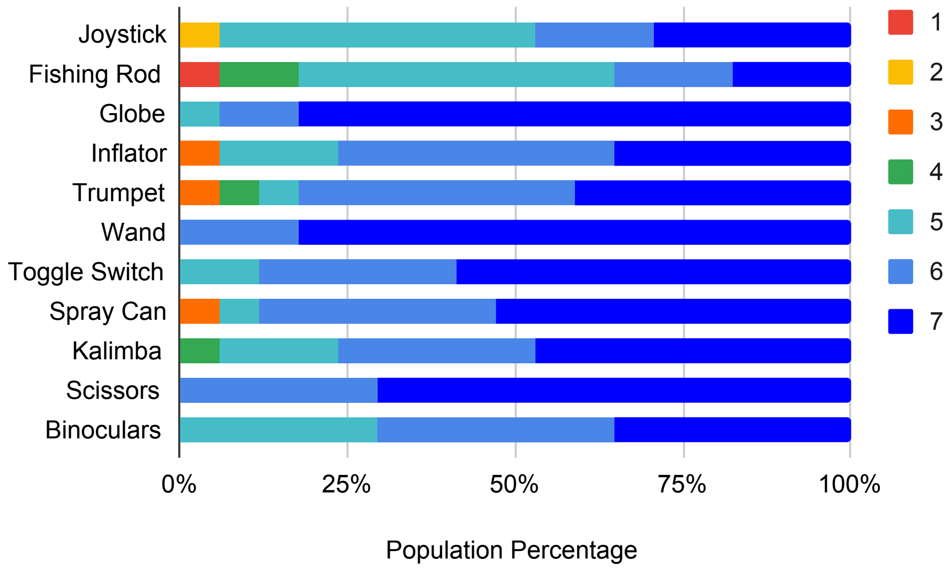

Figure A.27: Interactivity Evaluation of Hand Interfaces per Interface on Movement

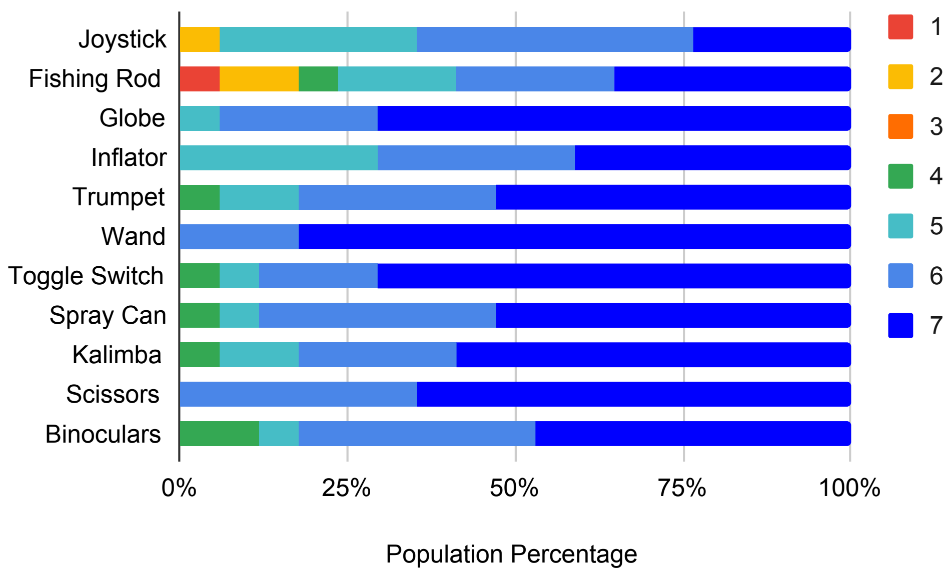

Figure A.28: Interactivity Evaluation of Hand Interfaces per Interface on Swiftness

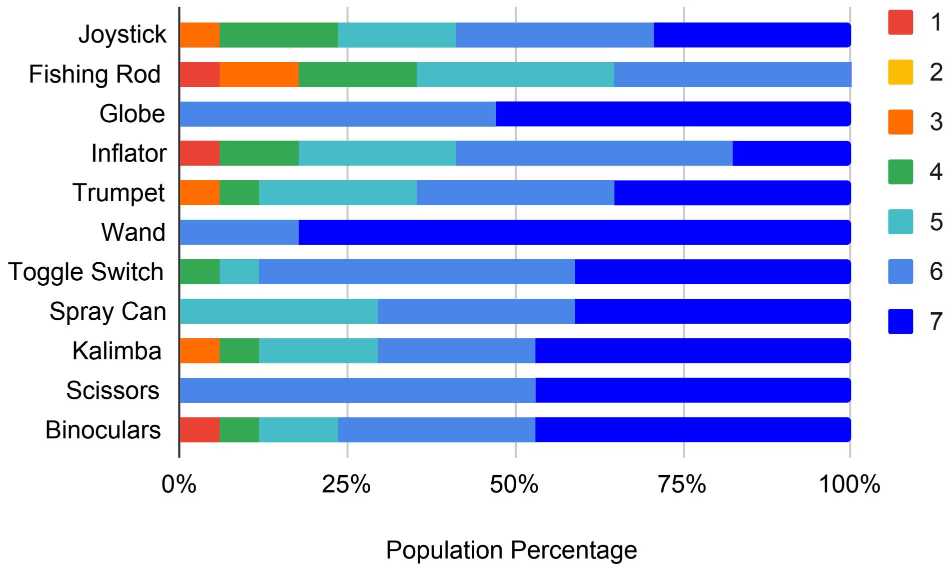

Figure A.29: Interactivity Evaluation of Hand Interfaces per Interface on Comfort

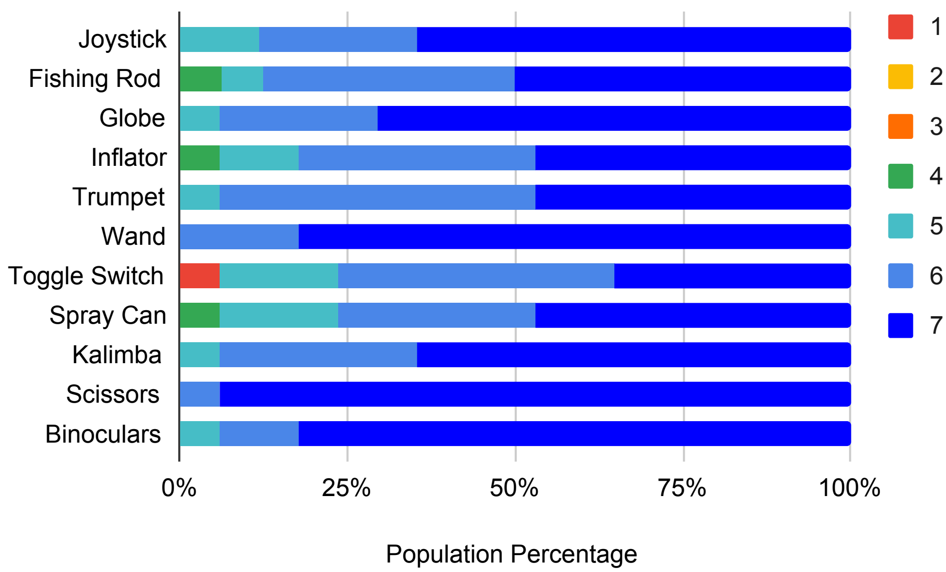

Figure A.30: Interactivity Evaluation of Hand Interfaces per Interface on Recollection

Bibliography

[1] 2017. The Principles of Good UX for Augmented Reality. https://uxdesign.cc/the-principles-of-good-user-experience-design-for-augmented-reality-d8e22777aabd Accessed August. 31, 2021.

[2] 2019. 18 AR, MR and VR Design Principles. https://www.justinmind.com/blog/vr-design/ Accessed August. 31, 2021.

[3] 2019. Hand Menu. https://docs.microsoft.com/en-us/windows/mixed-reality/design/hand-menu Accessed August. 31, 2021.

[4] 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/

[5] 2019. Introducing Instinctual Interactions. https://docs.microsoft.com/en-us/windows/mixed-reality/design/interaction-fundamentals Accessed August. 31, 2021.

[6] 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 Accessed August. 31, 2021.

[7] 2020. UX Design Principles for Augmented Reality. https://xd.adobe.com/ideas/principles/emerging-technology/ux-design-principles-for-augmented-reality/ Accessed August. 31, 2021.

[8] 2021. Designing for Hand Tracking. https://docs.ultraleap.com/xr-guidelines/designing-hand-tracking/ Accessed August. 31, 2021.

[9] 2021. Hand Physics Lab on Oculus Quest. https://www.oculus.com/experiences/quest/3392175350802835/?locale=en_US Accessed September. 5, 2021.

[10] 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*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 522. https://doi.org/10.1145/3290605.3300752

[11] 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.

[12] 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

[13] 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*, Gloria Mark, Susan R. Fussell, Cliff Lampe, m. c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, 1497–1507. https://doi.org/10.1145/3025453.3026030

[14] 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*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, 564. https://doi.org/10.1145/3173574.3174138

[15] 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*, Pejman Mirza-Babaei, Victoria McArthur, Vero Vanden Abeele, and Max Birk (Eds.). ACM, 374–385. https://doi.org/10.1145/3410404.3414260

[16] John 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)*. IEEE, 289–300.

[17] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. 2007. Applying HCI Principles to AR Systems Design. (2007).

[18] 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.

[19] 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*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, 1–7. https://doi.org/10.1145/3334480.3382888

[20] 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*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Mark Drucker (Eds.). ACM, 662:1–662:14. https://doi.org/10.1145/3411764.3445576

[21] Mar Gonzalez-Franco and Jaron Lanier. 2017. Model of illusions and virtual reality. *Frontiers in psychology* 8 (2017), 1125.

[22] 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*, Kiyoshi Kiyokawa, Frank Steinicke, Bruce H. Thomas, and Greg Welch (Eds.). IEEE Computer Society, 751–752. https://doi.org/10.1109/VR.2018.8446259

[23] 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*, Pierre Wellner and Ken Hinckley (Eds.). ACM, 3–12. https://doi.org/10.1145/1166253.1166257

[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*, Jeffrey S. Pierce, Maneesh Agrawala, and Scott R. Klemmer (Eds.). 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*, Elizabeth D. Mynatt, Don Schoner, Geraldine Fitzpatrick, Scott E. Hudson, W. Keith Edwards, and Tom Rodden (Eds.). 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*, François Guimbretière, Michael Bernstein, and Katharina Reinecke (Eds.). 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)*, Carolyn Penstein Rosé, Roberto Martínez Maldonado, Heinz Ulrich Hoppe, Rose Luckin, Manolis Mavrikis, Kaska Porayska-Pomsta, Bruce M. McLaren, and Benedict du Boulay (Eds.). Springer, 154–168. https://doi.org/10.1007/978-3-319-93843-1_12

[28] 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)*, Randall Shumaker (Ed.). Springer, 660–669. https://doi.org/10.1007/978-3-540-73335-5_71

[29] Kurtis Keveleigh. 2021. Hand Tracking - Mixed Reality Toolkit. https:

//docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/features/input/hand-tracking

[30] David Kim, Otmar Hilliges, Shahram Izadi, Alex D Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 167–176.

[31] 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

[32] 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*, David I. W. Levin, Fanny Chevalier, and Alec Jacobson (Eds.). Canadian Human-Computer Communications Society, 308–316. https://doi.org/10.20380/GI2020.31

[33] 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*, Gloria Mark, Susan R. Fussell, Cliff Lampe, m. c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, 1471–1482. https://doi.org/10.1145/3025453.3025600

[34] 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*, Regan L. Mandryk, Mark Han-

cock, Mark Perry, and Anna L. Cox (Eds.). ACM, 446. https://doi.org/10.1145/3173574.3174020

[35] 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*, Gudrun Klinker, Hideo Saito, and Tobias Höllerer (Eds.). IEEE Computer Society. https://doi.org/10.1109/ISMAR.2009.5336511

[36] 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

[37] 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

[38] Franziska Mueller, Dushyant Mehta, Oleksandr Sotnychenko, Srinath Sridhar, Dan Casas, and Christian Theobalt. 2017. Real-Time Hand Tracking under Occlusion from an Egocentric RGB-D Sensor. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society, 1163–1172. https://doi.org/10.1109/ICCV.2017.131

[39] 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 *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*. 189–198.

[40] 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, Jane Carrasco Chew and John A. Whiteside (Eds.). ACM, 249–256. https://doi.org/10.1145/97243.97281

[41] Siyou Pei, Alexander Chen, Jaewook Lee, and Yang Zhang. 2022. Hand Interfaces: Using Hands to Imitate Objects in AR/VR for Expressive Interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI 2022, New Orleans, LA, USA, May, 2022 (Unpublished manuscript)*.

[42] 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*, Robert J. Teather, Chris Joslin, Wolfgang Stuerzlinger, Pablo A. Figueroa, Yaoping Hu, Anil Ufuk Batmaz, Wonsook Lee, and Francisco R. Ortega (Eds.). ACM, 20:1–20:11. https://doi.org/10.1145/3385956.3418962

[43] 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*, David Kurlander, Marc H. Brown, and Ramana Rao (Eds.). ACM, 79–80. https://doi.org/10.1145/237091.237102

[44] 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

[45] 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

[46] 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

[47] 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*, Gloria Mark, Susan R. Fussell, Cliff Lampe, m. c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, 3891–3902. https://doi.org/10.1145/3025453.3026005

[48] 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.

[49] 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

[50] 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.

[51] 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*. 1–14.

[52] 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*, Hideki Koike, Carlo Ratti, Yuichiro Takeuchi, Kentaro Fukuchi, Stacey Scott, and Diego Martínez Plasencia (Eds.). ACM, 441–444. https://doi.org/10.1145/3279778.3279918

[53] Ultraleap. 2021. Leap Motion Controller. https://www.ultraleap.com/product/leap-motion-controller/

[54] 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

[55] 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*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, 42. https://doi.org/10.1145/3173574.3173616

[56] 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*, Re-

gan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, 78. https://doi.org/10.1145/3173574.3173652

[57] 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*, Beth Adelson, Susan T. Dumais, and Judith S. Olson (Eds.). ACM, 459–464. https://doi.org/10.1145/191666.191822

[58] 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*, François Guimbretière, Michael Bernstein, and Katharina Reinecke (Eds.). ACM, 1151–1159. https://doi.org/10.1145/3332165.3347869

[59] 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*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 111. https://doi.org/10.1145/3290605.3300341

[60] 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*, François Guimbretière, Michael Bernstein, and Katharina Reinecke (Eds.). ACM, 387–402. https://doi.org/10.1145/3332165.3347906