Affordances
for **Multi-device Gestural Interactions** in **Augmented Reality**

Shengzhi Wu

Affordances for Multi-device Gestural Interactions in Augmented Reality

Shengzhi Wu, Author

Molly Steenson, PhD, Primary Advisor

Daragh Byrne, PhD, Secondary Advisor

A thesis submitted to the School of Design, Carnegie Mellon University, for the degree of Master of Design in Interaction Design

ACKNOWLEDGMENT

ACKNOWLEDGMENT

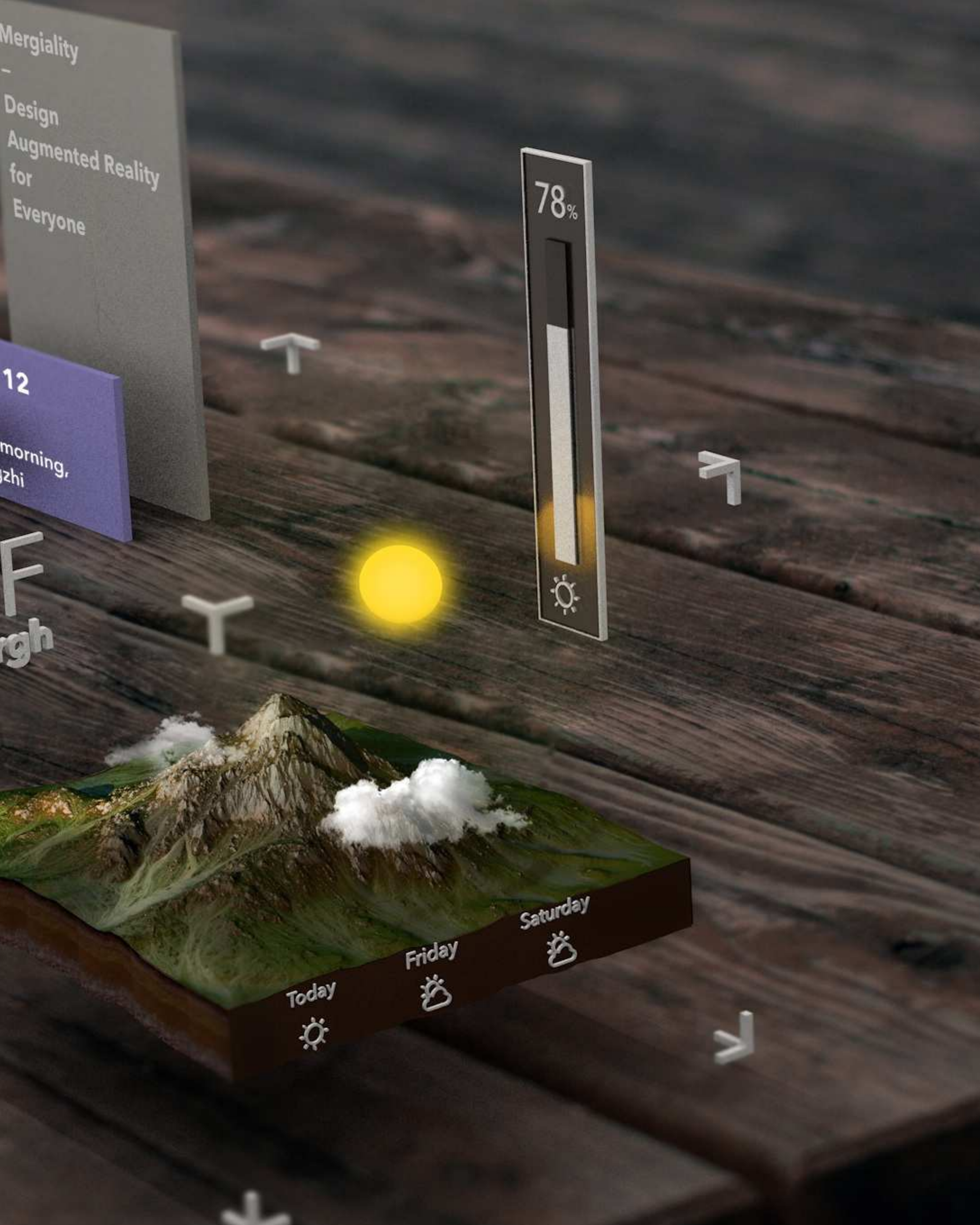
I express my sincerest thanks to Molly Steenson and Daragh Byrne, my thesis advisors, for supporting me when I was lack of resource, challenging me when I was reluctant to challenge myself, guiding me when I was unsure for the next steps, inspiring me when I ran out of ideas. Without them, I would never achieve so much for my thesis. More importantly, their passion, knowledge, and insightfulness set me a great role model for my future career and keeps prompting me to stay humble and curious.

I want to thank my wife, Tan Tianwen. She is always willing to lend her ears and share her opinions, although we were experiencing a long-distance relationship throughout the two years of my study. For me, she is more than a life partner, but also a great mentor. She can always point out things that I neglect and help me to think out of the box when making big decisions.

I also want to thank my parents for supporting me to pursue a master degree in the United States. Without them, I would not have the opportunity to further my study and finish this research. Their unconditional love makes me confident and brave enough to embrace all kinds of challenges in my lives.

In addition, I'd also like to thank all the people who participate in my research and share their valuable opinions to help me better iterate my design. Moreover, I want to thank Bruce Hanington and the School of Design for purchasing a Zed Mini camera, which allowed me to finish high-fidelity prototypes for my research. Also, I'd love to thank Chris Orris and Stereolabs for donating me another Zed Mini to support my future AR research.


Apirl
85°


Mergiality
–
Design
Augmented Reality
for
Everyone
12
morning,
zhi
78%
Today
Friday
Saturday

ABSTRACT

ABSTRACT

This thesis research is seeking to explore AR gestural interface design for multi-device controls (e.g., smart lights, smart speakers, computer displays, tablets, etc.) This design problem encounters three significant challenges. First, AR gestural interface design is still an emerging technology, with which most users are not familiar, it, therefore, requires the design must communicate itself intuitively even with novice users. Second, people have various mental models of interacting with different devices, so we need to keep their mental consistent by reducing their cognitive load. Third, gestural interfaces initially hold various constraints, such as easy to cause user fatigue, lack of tactile feedback, inaccurate tracking, etc.

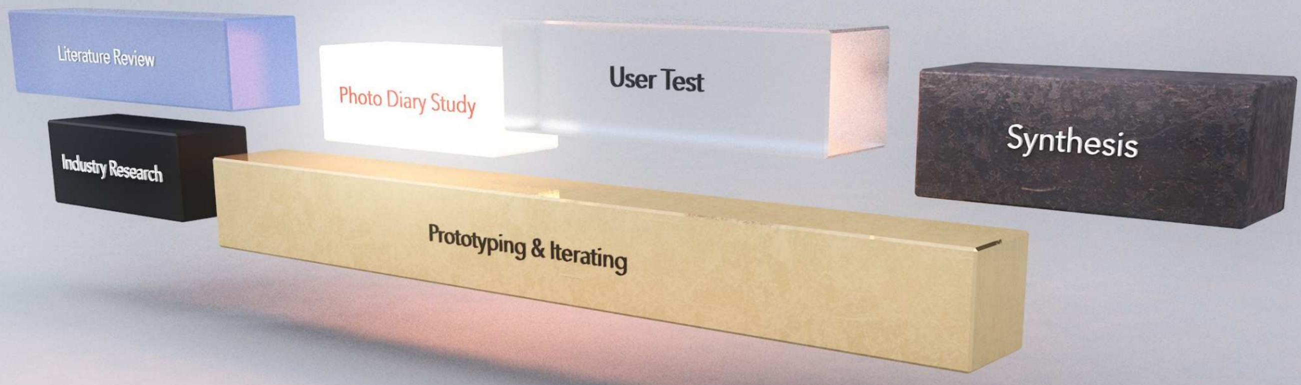
To address all those problems, this thesis research attempts to employ the theory of affordances into AR interaction design and explores the new feasible design paradigms of the future AR gestural interface design. The study mainly takes the approach of diary study, rapid prototyping, user testing and iterating, ultimately proposes three design paradigms of interactions that can inspire the future domain works.

CONTENT

CONTENT:

CHAPTER 1

FRAMING & LITERATURE REVIEW



Introduction

AR helps us to see information that is previously missing, hidden or non-existent. Nowadays, we can use a smartphone to control many smart devices through apps, but it's often quite cumbersome to use. Nevertheless, in the future where AR headsets are comfortable and light enough to wear, I believe that AR will become a new medium that can integrate with every type of devices seamlessly. We should be able to directly use our hand gestures to interact with multiple devices (e.g. a smart speaker, a tablet, computer displays, and smart lights) through the lens of AR. So the question is:

how might we design gestural interactions in AR to control multiple devices?

To explore these possible design paradigms, I started with literature reviews and industry research, aiming to learn the existing domain-works; then a photo diary study helps me to understand people's mental models of affordances from the physical world, which would also be applicable in AR interaction design. In addition, this research harnesses the method of "learning through making," and I utilize prototyping, iterating and testing as the major method to explore the design concepts. With two sets of experiments (basic interactions, and contextualization & application), I ultimately synthesis and propose five design recommendations, which, I believe, can benefit future work in this domain.

Figure 1.1: *Early AR headset*

Literature review

The following section reviews the three primary areas of literature behind the research, which includes the framing of AR, the theories of affordances, and the design of the gestural interface.

Framing of augmented reality

What AR means in this thesis

Augmented reality (AR) has existed for decades. In 1990, Boeing researcher Tom Caudell (Caudell, T. P., & Mizell, D. W., 1992) first coined the term "augmented reality" to describe a display that can blend virtual graphics onto physical reality. Then in 1994, Paul Milgram and Fumio Kishino (Milgram, P., & Kishino, F.,1994) used a different term, "mixed reality (MR)," which attempts to categorize different types of displays that involves the merging of real and virtual worlds. They also introduced the concept of virtuality continuum (VC) to describe the spectrum from real environment to a virtual environment. In Milgram and Kishino's (1994) classification, MR is the whole spectrum of VC except the real and virtual environment. In their definition and as depicted in Figure x, AR becomes a subcategory of mixed reality.

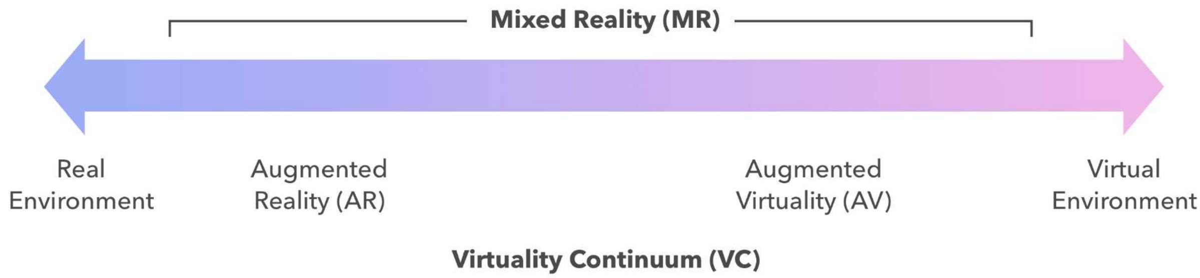


Figure 1.2: *mixed reality spectrum by Milgram and Fumio*

Figure 1.3: *Zed Mini pass-through AR camera*

Today, the primary distinction of mixed reality and augmented reality is also understood as different display techniques for headsets. Mixed reality is usually referred to as a camera passthrough solution for VR headsets. For example, Stereolabs frames their product as mixed reality, which is a Zed-mini camera on a VR headset that can pass the camera feed into users' eyes. On the other hand, an AR headset usually suggests headsets like Microsoft Hololens, or Magic Leap One, where the lens of the headset is transparent, and a user can see the actual environment.

This definition is very confusing when tangled with Milgram and Kishino's (1994) concept of mixed reality, and it fails to suggest any essential differences between the two concepts.

Figure 1.4: *Magic Leap One AR headset*

In this thesis, augmented reality is defined more broadly than a display technology, but an overarching term of technological augmentation of our reality. We must realize the concept of AR today goes far beyond visual displays, and it also includes spatial sound, environment localization, tactile and olfactory augmentation, etc. Therefore, the term AR best describes the augmentation we produced on our reality, whether it is visual, aural, olfactory, tactile, or gustatory. Additionally, I will concentrate on Head Mounted Displays (HMD) AR, which is promised to provide the best AR experience. And this thesis will not explore beyond visual and audio augmentation.

Applying theories of affordances

Bridging the gaps between physical product and AR design

Augmented reality exists as a parallel of our actual reality, and it merges itself into our physical environment. This thesis hypothesizes that theories of affordances that are previously widely used in physical product design, and human-computer interaction (HCI) could also be applied to AR design. By understanding the physical affordances and bridging insights into AR design, we could examine the commons and distinctions in between, and it helps us to better understand users' mental models. (better design not understand

The very first concept of affordances was put forward by James J. Gibson (Gibson, J. J.,1979) in his book *The Ecological Approach to Visual Perception* (1979), and it refers the "actionable properties of between the world and an actor (a person or animal)." While Donald Norman's (1988) book *The Psychology of Everyday Things* introduced the term affordances into the vocabulary of design and human-computer interaction, unlike Gibson's affordances, Norman proposes that affordance is the design aspect of an object which suggests how the object should be used (1998). This definition was soon taken up by the design and HCI communities, while the ambiguity of the definition simultaneously caused many confusion and misuses. Later, Norman (2008) realized the misleading nature of his original definition, and clarified the term "affordances" as "perceived affordances," as it emphasizes the significance of people's perception rather object's actual affordances, and he added "What the designer cares about is whether the user perceives that some action is possible". (Norman, 2008, p. 3), so this thesis will mainly use Norman's definition of perceived affordances as reference for discussion.

This clarification leads to some apparent differences between Norman's and Gibson's notions of affordances. As researchers later wrote, "Norman talks of both perceived and actual properties and implies that a perceived property may or may not be an actual property, but regardless, it is an affordance." (McGrenere & Ho, 2000) Although Norman's perceived affordances derived from Gibson's original

"Separating affordances from the information available about them allows the distinction among correct rejections and perceived, hidden and false affodances" (Gaver, 1991)

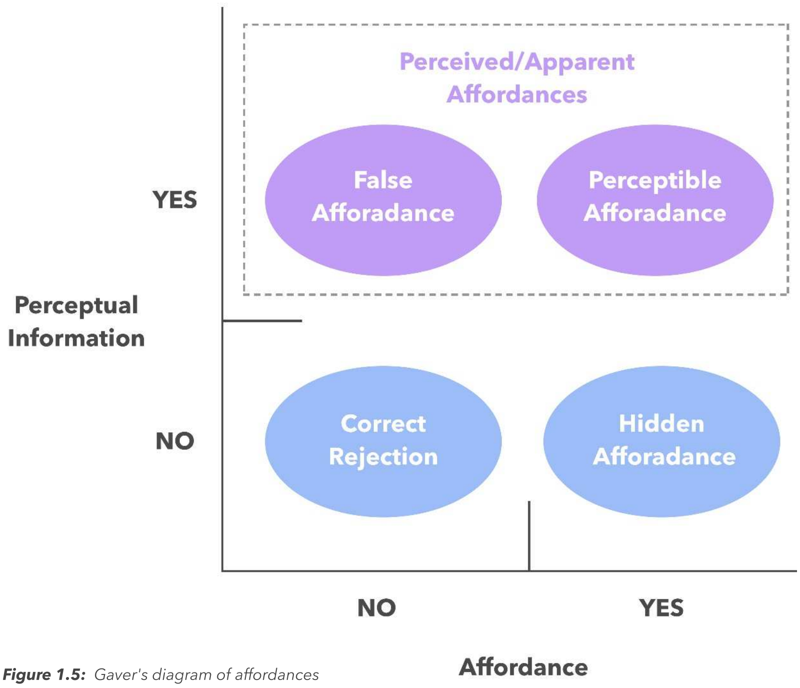


Figure 1.5: *Gaver's diagram of affordances*

concept, the clear distinction is that perceived affordances may or may not exist, and it purely determined by the people (users) who perceive them. As such, people's perceived affordances may vary vastly due to their knowledge, culture, ability to perceive and experience.

In addition, William Gaver's (1991) framework addresses the relationship between perceptual information and affordances explicitly. Distinguishing the relationship between affordances and perceptual information can be quite helpful for designers and clarify their goals. As Gaver points out, the commonly referred affordances are the "perceptual information available for an existing affordance,"(Gaver, 1991, p. 80) and he named it as perceptible affordances. On the contrary, the false affordances indicate that "information suggests a nonexistent affordance." While in most cases of design and HCI, we would want to reduce the false affordance and provide clear guidance to enhance the so-called perceptible affordance. Moreover, Gaver's framework also addresses the situation, where "no information available for existing affordances," (1991) and it's framed as a hidden affordance.

The question of affordances in AR serves as a fundamental question in this thesis, and some of the above-mentioned theories will be applied or discussed in the later sessions.

Gestural Interface Design

Gestural interface has drawn increasing attention in recent years, especially along with the popularity of virtual and augmented reality, and it holds the promise of offering natural and intuitive human-computer interactions.

Figure 1.6: *Google Soli Project*

To address gesture control for both large scale and small scale hand interactions, Barrett Ens et al. (2017) came up with a novel model of multi-scale gestural interaction for augmented reality, in which they combined Google Soli for precise micro-scale gesture recognition, and utilized Leap Motion for macro-scale recognition. As such, the gesture control can both provide accurate input and address the problems of the awkwardness and fatigue caused by direct manipulation. Nevertheless, I must admit that combining the Google Soli with Leap Motion is an extremely cumbersome setup, and still far from practical usage outside the research lab. This large-scale and small-scale interaction is related to the concept of *hand-space* and *world-space*, which will be discussed in later chapters.

The research of gesture controls does not only come from the academic field, attracted by the potential commercial use cases, and a few leading companies also take the initiative of designing better and intuitive gesture control schemes. Microsoft Hololens integrates a few gestures into its Mixed Reality Toolkit, such as air tap, bloom, which are quite limited; then, the recent released Hololens 2 supports more gestural interactions, such as pinch, grab, etc. However, Leap Motion, one of the most leading gesture recognition solution provider, conducted more in-depth research in the problem space. Its designers Eugene Krivoruchko and Keiichi Matsuda's (2018) project Cat Explorer had a deep thought on in-air hand-gesture interactions. Their article "Designing Cat Explorer" pointed out the challenges of the lack of tactile feedback when

Figure 1.7: *Leap Motion "Cat Explorer"*

designing in-air hand-gesture control, and it also raises problems of erratic motions. The article points out the importance of affordances in gestural interactions, and suggests to use pinch gesture because "Dynamic hand-aware affordances based on pinch gesture has allowed keeping the interface elements visually subtle without losing the ease of use and precision of control." (p. 11) In addition, they also proposed a handlebar UI combined with the pinch gesture for changing scales, and it offers stable controls and intuitive affordances.

Summary

The definition of augmented reality in this thesis is slightly apart from a particular display presentation or technique, but as a term to describe the overarching technological augmentation on our reality. Within this context, I strive to apply the theories of affordances into the AR interface design, bridging the gap between the knowledge we accumulate in product design and GUI design into the emerging field of AR interaction design. Using the theory of affordances in AR gestural interface, we should enhance the correct perceived affordances to communicate intuitive interactions, while avoiding the hidden or false affordances. Moreover, many considerations could draw from the precedents of gestural interface design, such as the lack of tactile feedback and erratic motions.

The following chapter will list several primary design challenges and considerations at the initial research stage. Those considerations also partially derived from the literature reviews and will be used to guide the overall design research. Also, the technologies that are used for prototyping will also be briefly discussed in the next chapter.

2

CHAPTER

DESIGN CHALLENGES, CONSIDERATIONS & TECHNOLOGIES

DESIGN CHALLENGES, CONSIDERATIONS, AND TECHNOLOGIES

In the problem space of AR gestural interactions and multi-device controls, there are many existing design challenges to acknowledge. Those challenges become later design considerations in this thesis. To address these problems, my major approach was to research through design, and design through prototyping. As such, in this chapter, I will also discuss technologies involved throughout the process, and shed light on the recommendations of prototyping techniques for future AR designers.

Figure 2.1: gestural controls in movie *Minority Report*

Challenges and Considerations

During my initial research, I listed a few critical high-level design challenges and design considerations that either I am encountering or will be encountering, also problems that precedents discussed in their research. I needed to be aware of these challenges and considerations, and seriously consider or elegantly address them throughout my research. These design consideration prompted and guided the project.

1. Transfer the screen-based experience into immersive computing

Considering the transition of 2D screen-based interface to 3D immersive computing can be a double-edged sword. On the one hand, utilizing people's existing mental models of screen-based interaction can build a relation to their existing familiarity and reduce the cognitive load for many novice users. On the other hand, directly placing 2D elements into the immersive computing space may neglect many valuable opportunities to leverage users' embodiment and three-dimensional space sufficiently.

2. User's comfort and ergonomics

Gestural interactions in science fiction movies, like in the movie Minority Report (Spielberg, S. et al., 2003), looks fascinating, but it fails to consider ergonomics and put a user in the center of the design. We need to realize that people usually get fatigued very quickly when holding up their hands in the air, therefore when designing hand-gestural interactions, we must consider users' comfort level, and the ergonomics

Figure 2.2: Ideal gestural interaction height range for sitting

of their hands, especially certain gestures are more fatiguing than others. The position and orientation of the interactive elements should also be carefully considered to make them easy to access. Ergonomics and user's comfort aren't a new question, Leap Motion article "Ergonomics in VR Design" (Leap Motion, 2016) discussed many common issues and recommendations for how to address them , such as ideal height range, restrict users' motions to reduce their motion sickness.

3. The lack of tactile feedback

A significant challenge in designing freehand-gestural interactions is the lack of tactile feedback. Designers Eugene Krivoruchko and Keiichi Matsuda (2018), who creates Leap Motion Cat Explorer, suggest the lack of tactile perception and material support as the most crucial issue. In the tangible interactions, interacting with physical objects, such as a button, a dial or a slider, always associate with tactile feedback. Even very subtle tactile feedback can still provide a strong signal, making a user feel confident and concrete about the actions he or she performs. As such, other additional feedbacks, such as visual, audio, need be provided to compensate for it.

4. Erratic motion and accidental trigger

Since gesture recognition technology is still far from perfect, while designing the gesture controls, designers have to consider that the gestures need to be not only intuitive for a user to understand, but also accurate for a machine to recognize correctly. Additionally, designers need to avoid gestures that may be triggered accidentally or unintentionally. Furthermore, human's hands are inherently erratic, especially when we have our hands in the air; therefore the raw gesture recognition data might be jittery and pick up erratic motions. Thus, minimizing users' awareness of the erratic motion in the interactions also becomes an interesting design challenge.

5. The intuitiveness of the hand-gesture

Hand-gestures have been used for thousands of years as a vital part of body language. They already form certain conventions in term of the meanings of different gestures; thus designers should be aware of that the semantic of gestures may vary in different countries and cultures. Furthermore, we need to think about the affordances of gestures, understand how people interact with physical objects, and consider how to bring people's existing mental models of gestures into the design of AR interfaces.

6. Seamless transition between devices

Working with multiple devices can be cumbersome, and will burden users' mental and cognitive load when switching from one device to the other. When designing a user's experience of using multiple devices, we should try to make the technology itself blend into the background, placing the task as the center of the focus instead of the tools. Therefore the interaction should be able to ease the transition and keep users' mental models consistent from one platform to another to make the experience seamlessly.

The above-mentioned points are the initial challenges and considerations, and those high-level considerations will guide the project throughout the whole process. While on the other hand, other considerations do exist, such as addressing accessibility issues for both right-hand and left-hand users, or those who are not able to use hands in a fully functional manner, and the social awkwardness when using it in a public space. However, those later mentioned considerations are not placed as a primary focus for this particular project.

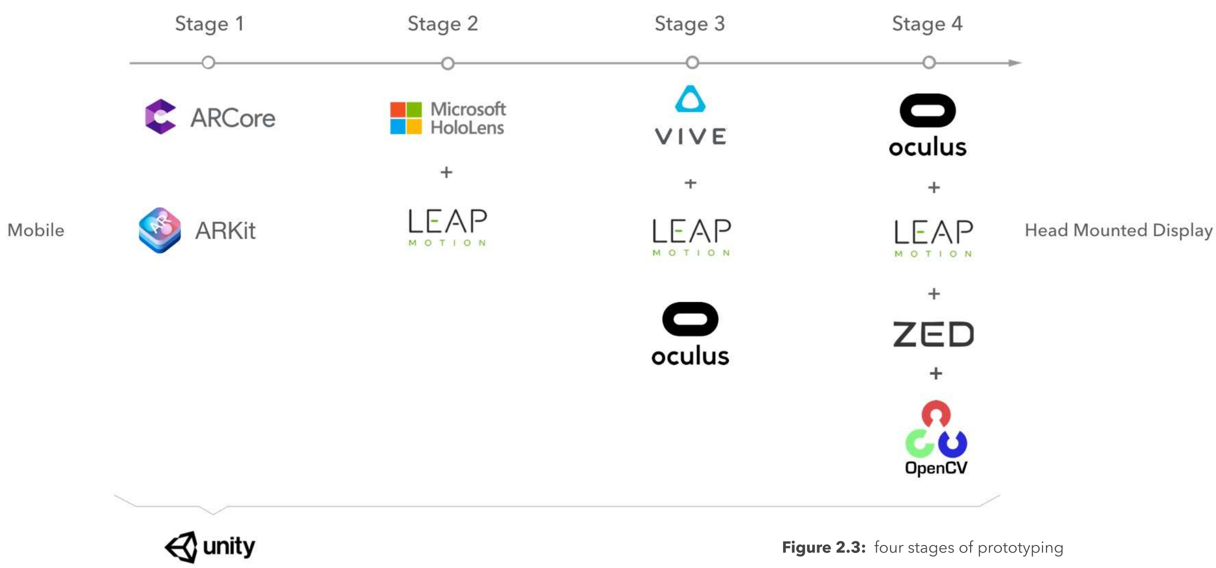


Figure 2.3: four stages of prototyping

Prototyping Techniques

The prototyping process contains four major stages, which are hand-held mobile AR, Hololens + Leap Motion, Oculus Rift + Leap Motion and the final stage with Rift, Leap Motion, Zed Mini and OpenCV.

Figure 2.4: Google ARCore and Apple ARKit for mobile AR platforms

1. Hand-held mobile AR

In the first stage, I mainly examined hand-held mobile AR platforms, using Google ARCore, Apple ARKit, and Vuforia image marker for mobile prototyping. The mobile AR SDK is the most approachable approach, and it assists designers to get familiar with current AR technology quickly. While the mobile AR holds many drawbacks, for example, it requires users to spare one hand to hold the devices, which significantly eliminate more complex and immersive interactions. Since this thesis mainly researches gestural interactions, so it only focuses more on head-mounted display (HMD) AR.

2. Hololens + Leap Motion

In the second stage, I moved forward to a HMD AR with Hololens, and placing Leap Motion on top of the Hololens to access the gesture input. However, the limited field of view (FoV) and the frequently occurred Unity crash problem impede me from further exploration. Especially since I used Hololens Remoting to stream the camera input back to my computer through Wifi network, it becomes challenging to have a stable tracking and record the demo video. This problem is presumed to be a Unity compatibility issue or slow Wi-fi bandwidth, and it may be addressed with a different version of Unity or with a better signal.

3. Oculus Rift + Leap Motion

I then started to use Oculus Rift mounted with Leap Motion to prototype the design concept. Since Leap Motion already provides fully functional example code, documentation, and the compatible SDK, the prototyping process becomes much smoother and faster. The tradeoff of using this pipeline is that I have to accept the fact that all the prototypes exist in a virtual reality environment rather than in augmented reality. Admittedly, virtual reality as one approach of spatial computing shares a lot of common interactions with augmented reality; therefore it is still possible to gain some insights from prototyping the concepts in VR.

Figure 2.5: Oculus Rift + Leap Motion + Zed Mini setup

4. Oculus Rift + Leap Motion + Zed Mini + OpenCV

At the final stage, I combined Oculus Rift, Leap Motion and Zed Mini (a passthrough AR camera), and eventually reach the result of stable tracking in AR. This is a highly recommendable pipeline for designers who want to execute high fidelity rapid prototyping. However, Vuforia is not compatible with this pipeline; thus in order to achieve object tracking, I ended up using OpenCV and ARUco for image marker tracking. Other commonly used technologies are also helpful in this development environment, such as Open Sound Control (OSC) and Unity Network for communicating information between ifferent devices, Philips Hue as smart light devices, and Google Home as a smart speaker.

Reflection

Designing gestural interactions in AR requires designers to put human in the center of the design, by thinking about the intuitiveness of gestures, users' comfort, and ergonomics as a premise. Moreover, due to the lack of material support, gestures in the air cannot reply on any tactile feedback, which will likely cause erratic motions and inaccurate tracking. What's more, controlling multiple devices adds another layer of complexity on users' cognition. Certainly, there're more issues in this domain, but these aforementioned issues are directly related to the topics of this thesis and will be addressed and discussed in later chapters.

Also, since this research utilizes prototyping and iterating as main methods for design research, so I explored and compared various available technologies. With those prototyping techniques, I created 24 working prototypes with over 10,000 lines of C# codes to experiment and test different design concepts.

3

CHAPTER

INITIAL RESEARCH
–
PHOTO DIARY STUDY & INDUSTRIAL RESEARCH

INITIAL RESEARCH: PHOTO DIARY STUDY AND INDUSTRY RESEARCH

The initial research allowed me to gain a fundamental understanding of the problem space. I conducted a photo diary study as user inquiry to gain insights on people's mental models of affordances, and also researched the existing design paradigms of AR gestural interface from industry. The insights of both help me to identify a framework of five critical elements that affect a user's perceived affordances, and the framework lays a foundation for later design concepts.

Photo Diary Study

Immersive computing (such as AR and VR) initially shares many commons interaction patterns with the physical world, and they both deal with 3D space and substantially leverage our body and muscle memories to engage with our environment. Thus it is probably reasonable to assume that we could gain a better understanding of users' mental models and the perceived affordances for augmented reality by digging into the physical interactions. Therefore, I conducted a photo diary study to understand people's mental models on physical products in our daily lives.

The photo diary study recruited six participants and prompted them to record photos of the products and objects in their everyday, particular the products and objects that seem intuitive to use, or those that are confusing to use. The goal was to extract the common patterns of affordances from collected photos.

The participants consist of students from the School of Design, Entertainment Technology Center (ETC) and Heinz College of Carnegie Mellon University. The selected participants are composed of equally female and male, and subjects were not required to have a design background or knowledge.

After using affinity diagramming to map out the similar concepts from 60 photos collected by the participants within three weeks, I concluded the following key insights from the activity:

- *Ergonomic shape reveals the affordances to users. For instance, a mouse in Figure 2.1 can fit a person's hand very well, and it invites him to put his hand on it.*

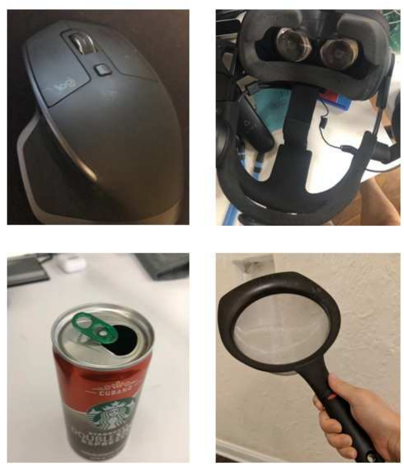


Figure 3.1: *Ergonomic shape reveals the affordances*

Figure 3.2: *Positive and negative shapes disclose their connects*

- *Signifier (text, icon e.g.) can amplify the perceived affordances, especially indicating products' functionalities. For example, a microphone icon with a crossing line clearly communicates its function.*

- *Positive and negative shapes that match each other disclose their connections, making the affordances perceived to be clear.*

- *Lights can indicate the status of the system, also provide directional guidance.*

The key insights later informed the design concepts and decisions, especially in applying the physical affordances into AR interactions. The implementation will be discussed in Chapters 4 & 5.

Figure 3.3: *Light is used to provide directional guidance.*

Industry research

The augmented reality industry is evolving rapidly; therefore I looked at the AR interaction and interface design from existing products, such as "Cat Explorer" from Leap Motion, Hololens gesture-based game RoboRaid and so on. Additionally, I read the design guidelines from Magic Leap, Google, Apple, and Windows mixed reality, and it enables me to understand the existing design paradigms from the major players in the industry.

Moreover, I also examined the demo prototypes that designers and developers posted online to keep the trend with the rapid-moving industry. For example, AR/VR developer Yujin Ariza and his team at ETC created Project Pupil, which uses hand gestures as the primary input for a multi-person learning experience. I also talked with Martin Schubert, who is formally a designer at Leap Motion, now at Oculus, and discussed his project of gesture shortcut; additional, I also met Noah Zucker, who did many inspiring explorations in VR gestures controls and posted them on Twitter. Those are only a tiny part of the examples I inspected during the industry research. I played with over 40 different AR/VR products and deeply investigated about 35 example demos online to gain a deep understanding of the industry.

Figure 3.4: *Leap Motion "Project North Star"*

Based on the industry research and diary study, I developed an affinity diagram with the principal elements, and common patterns appear, which allows me to map them into five key categories, including shape, lighting, color & shading, haptic & acoustic, and motion. An example may help to illustrate the patterns. In the Leap Motion "Cat Explorer" (2018), a circular shape of the base indicates its affordances of rotation. While a user's hand approaches it, the part that is close to his hand brightened, and it uses color and lighting to provide a proximity visual feedback; then when a user touches it, a subtle sound effect builds up the experience; eventually when a user moves his hand accordingly, the base moves along with it simultaneously. In this small example, color, light, sound, and motion combine for communicating clear affordances. Similar patterns occur repeatedly, and some use one or two of them, others may use more at the same time. Those five elements significantly affect how we perceive interactive content, as well as the affordances it communicates. Later with my thesis research, I use this framework to guide on how to deliver correct affordances in AR interfaces.

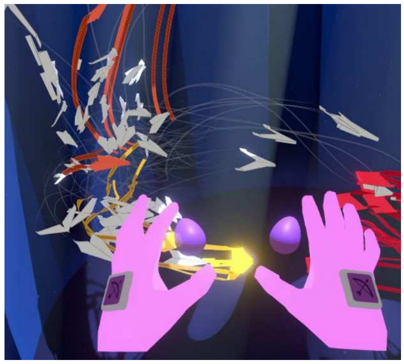


Figure 3.5: *Martin Schubert's project "Shortcuts" at Leap Motion*

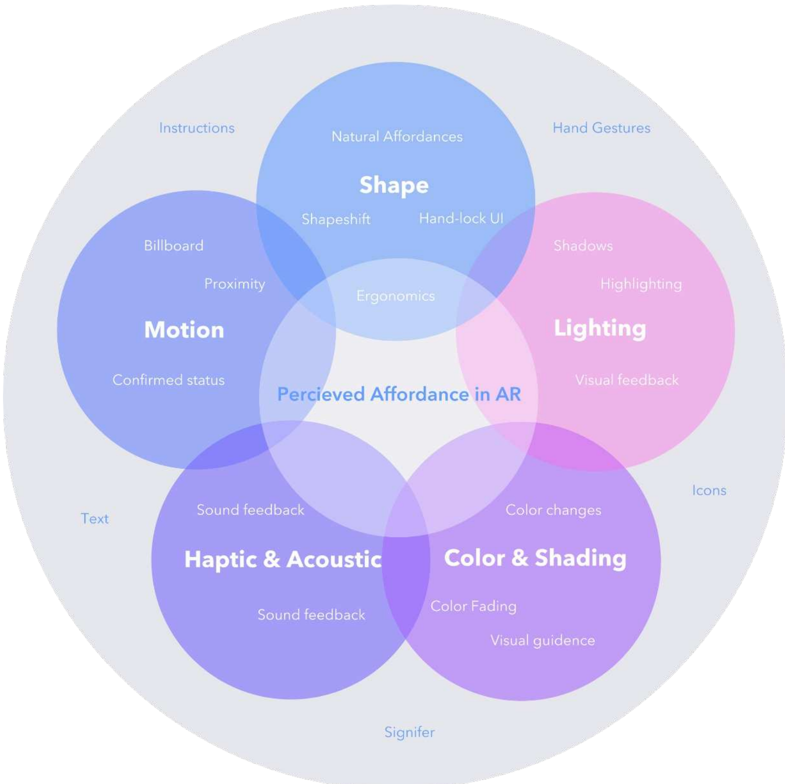
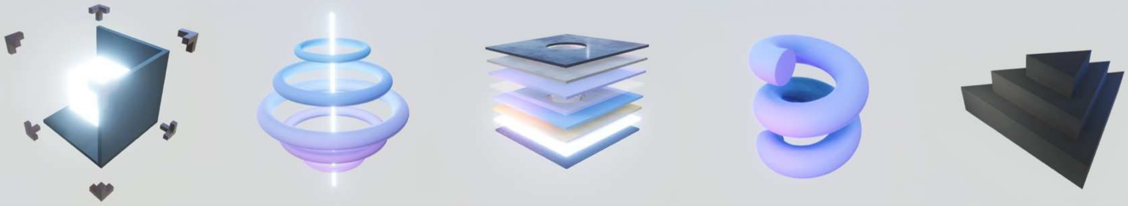


Figure 3.6: *Affinity diagram of the key components from diary study and industry research, and the final result leads to a five-element framework.*



Lighting Motion Color & Shading Acoustic & Haptic Shape

A framework of manipulating perceived affordance

As discussed earlier, common patterns of the diary study and industry research disclose five key elements that affect people’s perceived affordances: shape, lighting, motion, color and shading, and acoustic and haptic. By manipulating the attributes of those five elements, we can enhance a user's perceived affordances, either revealing the hidden affordances or enhance the existing perceived affordances and give users more solid feedback.

Lighting

Lighting itself is not obviously visible in many situations, but it shapes the appearance of an object. Aligning the virtual lighting in AR with the actual environmental lighting creates a realistic and believable illusion, yet lighting can do more. In the traditional GUI system, the changes in lighting usually reveal as a modification of color, such as changing brightness or opacity of a color. While in AR interfaces, lighting could work with color and shading together to signal a system status or actions, such as providing proximity feedback. Also, light sometimes is used as a volumetric ray to point at specific directions, as the diary study shows that light can provide directional guidance. Moreover, lighting also helps to create a hierarchy among multiple objects, highlighting a specific object or information.

Motion

Motion brings attention to a particular object, and it serves a variety of purposes. Alike the hover state in GUI, motions, such as changing scales or positions, can signal a particular interactive object over others, making the UI feel responsive. Also, motion could simulate real-world physics in the virtual space. Especially the illusion of physics in the virtual space can

help a user align unfamiliar AR interactions with his physical experience. At last, motion can facilitate smooth transitions, and help a user to understand the changes from one state to the other.
Light can function as a ray for directional guidance

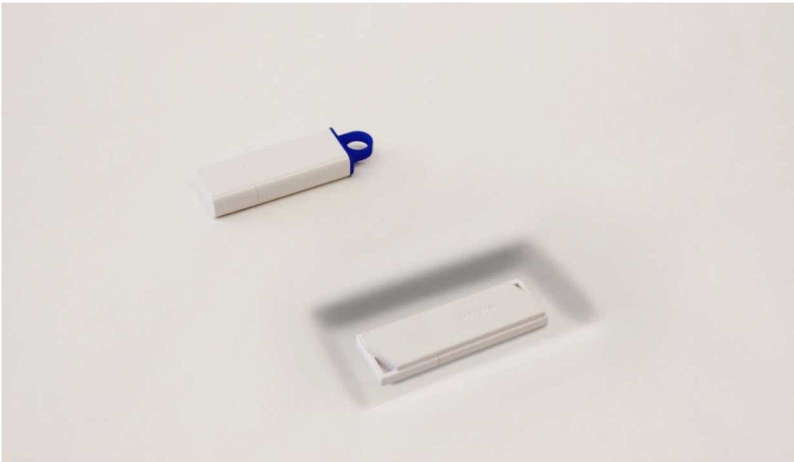


Figure 3.7: Dixon Lo's research uses shadow to sink a USB drive into a table to indicate the amount of data in it.

Color & Shading

Color and shading usually work with lightings, and together they determine how an object looks and communicate a particular status of an object. Similar to GUI design, we can take advantage of the co-relationships between colors and meanings that already exist in people's mental models. For example, a red color usually conveys the meaning of warning, danger, attention, etc.

Shading in AR could modify users' perception as well. With proper shading and shadowing, the perceived shape of an object can even change. Dixon Lo et al's (2018) research shows how perceived affordances and the meanings of an object are modified through the manipulation of shading and shadowing. For example, using shading and shadows to sink a USB drive into a table may indicate the weight and the amount of data inside the drive.

Haptic & Acoustic

Haptic feedback does not usually exist in gestural interactions without extra accessories, like haptic gloves, but it indeed plays an important role both in our physical interactions and in AR. Tactile feedback allows us to grab objects without looking at them, and we could even sense the shape of an object by merely touching it. Admittedly, the absence of haptic sometimes sabotages the immersion in AR, but it does not necessarily mean we should overlook it. Later research shows that encouraging a user's fingers to tap on each other can partially resolve the issue, and more discussion will take place in the later chapters.

Figure 3.8: a big shape button usually invites people to push with a full palm

Additionally, acoustic feedback, such as sound effects, may help to compensate for the deficiency of the lack of tactile and haptic feedback. We could think about how a haptic feeling can be transferred into a sound effect that still retains its quality. To retain the quality, we do not need to create precisely accurate sounds. For instance, we can barely hear any sound when we grab a physical object but triggering a subtle tap sound in AR as grabbing feedback could still generate a believable illusion.

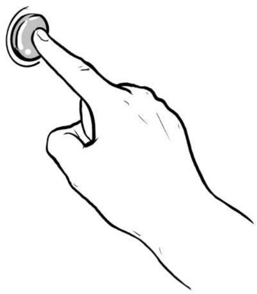


Figure 3.9: a small button usually encourages a user to use only one finger to press it lightly

Furthermore, 3D spatial audio can enrich the AR experience significantly. It can draw attention to a direction or an object outside a user's field of view, providing another layer of information that is subtle and nuanced. Ultimately, when designing the sound effect itself, we need to consider how it can fit into various kinds of environments. For example, dropping AR furniture on different kinds of surfaces may sound different. Therefore disregarding users' unpredictable environments, we could still try to design a sound effect that is generally applicable for most of the situations. However, sound and acoustic will not be the main focus of this resear

Shape

The shape of an object determines the way of how a user initially perceives it. It works similarly with the affordances in our physical world, since virtual content and real-world objects mostly look quite similar. It is reasonable to assume that users already build mental models of interacting with physical objects, and those mental models, like our muscle memories, immediately tells us what it affords by its shape. As found in the diary study, ergonomic shape reveals its affordances to users, so when designing for gesture interaction, we must consider how the shape affords a user's hand to interact with it. For example, a big shape button usually invites people to push with a full palm, and it might feel heavy; while a small button usually encourages a user to use only one finger to press it lightly.

Conclusion:

In this chapter, I proposed a framework to demonstrate how five elements affect a user's perceived affordances, and most of the time designers need to consider them together. It is the combination of them, rather than an individual element, defines how a user perceives it. The next chapters will implement these elements into contexts and interactions, and discuss when and how each of them would be useful.

4

PROTOTYPING AND ITERATIONS

–

BASIC INTERACTIONS

CHAPTER

Prototyping and Iterating I: Basic Interactions

The initial prototyping process starts from exploring basic interactions, including point, press, pull and slide.They are the most commonly used interactions that I identified through my industry research and diary study. The insights of these basic interactions will lay a foundation for more complex gesture interactions in the future study to afford applications in various contexts. In this chapter, I will discuss the process, iterations, and thoughts behind each interaction design. Not all of the interactions are successful, but it is those failures that are even more valuable for gaining in-depth insights, which will later turn into high-level design recommendations driving future research.

Interaction 1
Point with an index finger

Pointing with a thumb is one of the most natural and frequent-used gestures in our daily lives. This concept brigs the pointing interaction into AR. Although the gesture is proved to be intuitive to most users, it also has a few drawbacks hindering it from being an effective interaction for AR gestural interface.

Figure 4.1: *prototype of pointing with an index finger in VR*

First, a little minor inaccuracy of the hand tracking would make the pointing feels unsatisfying. Because the pointing direction is calculated with the connection f index finger joints, so when shooting a ray, a small deviation will enlarge over the long distance. As a result, a little jittering of hand tacking can cause dramatical erratic motions when it reaches to a target object.

Second, this gesture is useful when combining with another gesture for selection, but while a user is performing this gesture, any movement of other fingers will affect the orientation of our index finger, and it becomes frustrated to select a target accurately. That being said, this issue may be addressed with multi-modal interaction, such as using

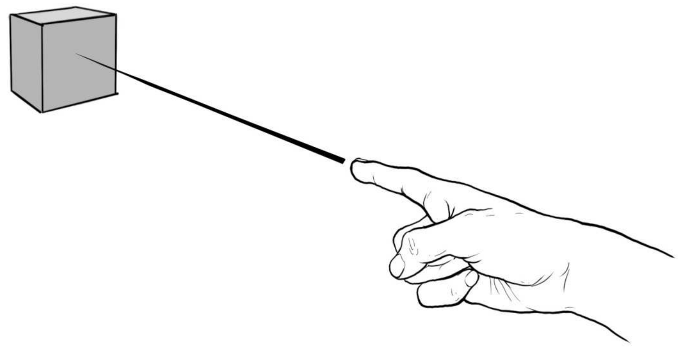


Figure 4.2: *pointing with an index finger is one of the most natural and frequent-used gestures*

voice and the pointing gesture together.

Finally, we must realize that this gesture can easily cause a user's hand to become fatigued over time, which is a major drawback and make it less usable for frequent or long-time interactions. This finding draws attention to the issue of users' comfort in gestural interface design. It is easy to neglect the fact that certain gestures naturally are more fatiguing than others. That is not to say we have to give up those gestures, but to evaluate the usage frequency, intuitiveness and tiredness of the gesture before making a design decision.

Interaction 2
Press

Press is commonly used for buttons or other types of selections. We can mostly reuse the principles we learned from GUI design for designing this interaction. One of the differences in AR interaction is that the mouse hovers state trigger will be replaced as hand/finger proximity. Then we can apply the "five elements framework" to modify each attributes to communicate visual or audio cues for the proximity feedback, such as changing color or shading.

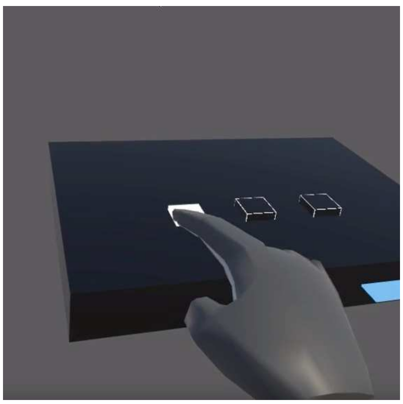


Figure 4.3*: press prototype*

Another significant distinction is the absence of tactile and haptic feedback. A user presses a virtual button in the air. In this case, realistic motion (e.g. a button pushes down when pressed and pop-up while released) and audio effect can enrich the experience greatly. Moreover, we can place a virtual button on top of solid surfaces, so that the tap on the surface will create tactile feedback to deceive a user's perception.

Interaction 3
Pull

The pull interaction initially comes from a physical chest of drawers. A handle or a notch communicates clear affordances of pulling. Often, the shape is ergonomically designed for our hands, and as my diary study shows that the ergonomic shape reveals the affordances. We can also leverage it into an AR interface design.

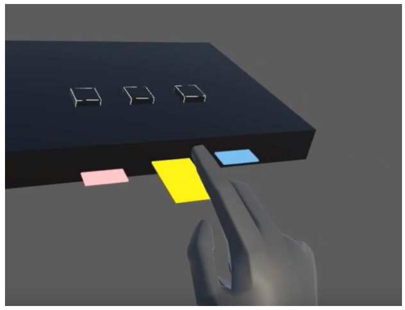


Figure 4.4: *proximity feedback of pull interaction*

In the meantime, the handle also evolves its digital forms in GUI. One favorite example is the menu bar at the bottom of iPhone X. By dragging or pulling it upwards, it takes users back to the home menu. We could still leverage the affordances that people learn through a digital interface to inform AR interaction design.

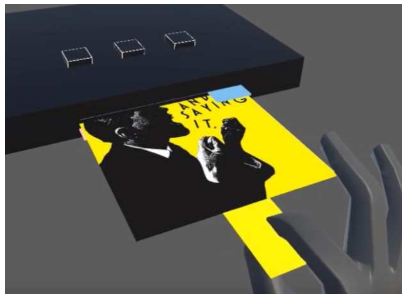


Figure 4.5: *pulling out content*

The initial design concept harnesses the familiarity from pulling a bookmark. For proximity feedback, the length of the handlebar extends when a user's hand gets close. Yet in another iteration, to make the handlebar visually subtle and less obtrusive, I designed a small rectangle bar sitting along with the interactive object, just like the handlebar at the bottom of the iPhone X.

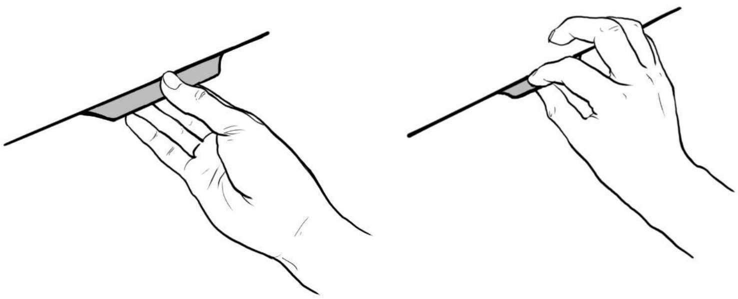


Figure 4.6: *the size of the handlebar affects how a user grab it*

An interesting finding from this interaction is that the size of the handlebar affects how a user grasps it. Users usually consider a more extensive or longer handlebar requires a full hand to grasp, while a shorter one fits better with a finger size, which allows them to grab by thumb and index finger with a pinch gesture. While technically speaking, a pinch gesture is easier to be detected accurately, so encouraging users to use a pinch gesture to grab and pull might be another design consideration.

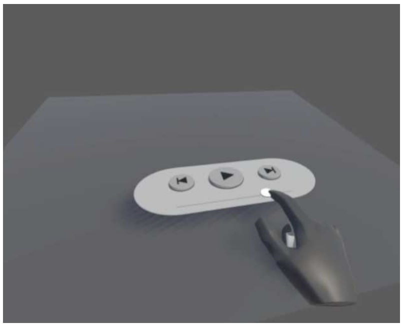


Figure 4.7: *a virtual slider that mimics real-world physical slider*

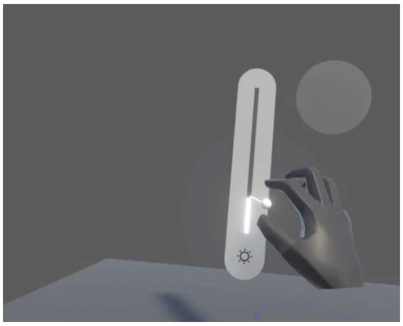


Figure 4.8: *pinch slider uses a small handle in between index finger and thumb for proximity feedback*

Interaction 4
Slide

Sliding affords adjusting the volume, such as modifying the brightness of a light. The first intuition is to mimic a physical slider into AR directly. However, soon after prototyping, it is very noticeable that the interaction does not feel satisfying due to the lack of haptic feedback. When we interact with a physical slider, the slider button offers a place to support a user's fingers, and a resisting force eases the fingers movement allowing accurate modification. On the other hand, simulating the exact same interaction in AR means that a user is holding up his or her hand in the air. Without a place to rest his or her hand, more erratic motion of the hand makes it tough to control precisely. Additionally, users feel less assured about the interaction due to the lack of tactile feedback.

As a result, the slide interaction is then inspired by Leap Motion "Cat Explorer" (2018) and uses a pinch slider to address those issues. The pinch slider firstly offers proximity feedback as a user's finger approaches, and a small sphere (handle) will appear between his thumb and index finger to encourage them tapping on each other and squeeze it. Once a user executes a pinch gesture, the slider bar will follow the user's pinch point in a single axis (usually Y-axis) with the ease-in transition until the pinch is released. The pinch interaction provides additional haptic feedback by encouraging a user's finger to tap on each other, and it makes the interaction feel more solid and concrete. Also, the ease-in transition reduces the effect of fingers' erratic motion, making it more precise and easy to control.

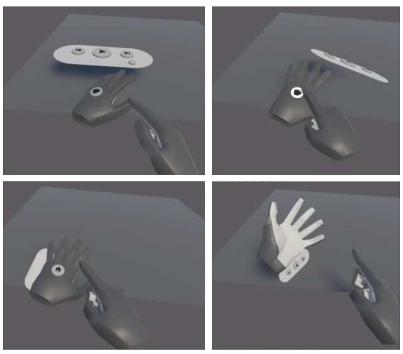


Figure 4.9: *sequence of transitioning hand view to world view*

Interaction 5
Transition from hand view to world view

In AR interactions, sometimes a user would like to have a shortcut to access frequently used functions, such as navigating back to the home menu or playing and pausing a piece of music or video. While in other situations, the interactive UI needs to fix to a specific place in the room. It either builds the relationship with a particular physical device and environment, or it helps a user to remember where it is. Therefore, allowing the UI panel to switch between hand view to world-view can create a seamless transition between the two modes.

This concept enables the mode switch between world-view and hand-view. When a user's hand gets close to a surface, such as a table, a button on the back of his hand will appear to indicate the available state. By using the other hand to tap on the button, a user can switch the views back and forth. One a strategy to place the UI panel is to find the closest point from a user's hand to the available surface.

This design evolves a new concept of world-view and hand-view, or we could also call it world-locked and hand-locked. Hand-lock means locking something to our hand, like a watch, or a ring we are wearing; whereas world-locked is the opposite, securing it to the actual environment surround us. In the last chapter, there is a detailed description of the two terminologies.

For the world-view, it helps a user to mentally build relationships of the UI and its physical environment or devices. Fitts's Law can partly be applied here, where a closer distance between a virtual UI element and a physical device reveals their relationship, and it is reasonable to assume that they are connected. On the contrary, placing a UI panel far from its controlled device may cause some confusion.

Hand-view allows a more close-up interaction, and it will move along with our hands, following us wherever we go. Since it becomes a part of our bodies, so it is usually easy and quick to access. But it should only display when necessary. For example, we could show it when a user's palm is facing his head. To better understand these two concepts, we could use as an analogy the hand-lock to the dock on a desktop system. No matter how we move application windows, the dock always stays at the bottom (depends on a user's preference) allowing quick access. In the later chapter, there will be more discussion and context of the two modes of interactions.

Interaction 6
Transition over distance

The gestural user interface is usually designed for arm-reachable distance, thus we need to consider the accessibility when a user intends to control a device in a distance. A straightforward solution is to transition the interactive UI close to the user when it's needed.

Combining the eye gaze with the hand orientation might be a solution. While a user is looking at a particular device, and his hand is facing the same direction, the UI will transition right in front of the user's hand. The interactive UI stays until either the user looks at a different direction or

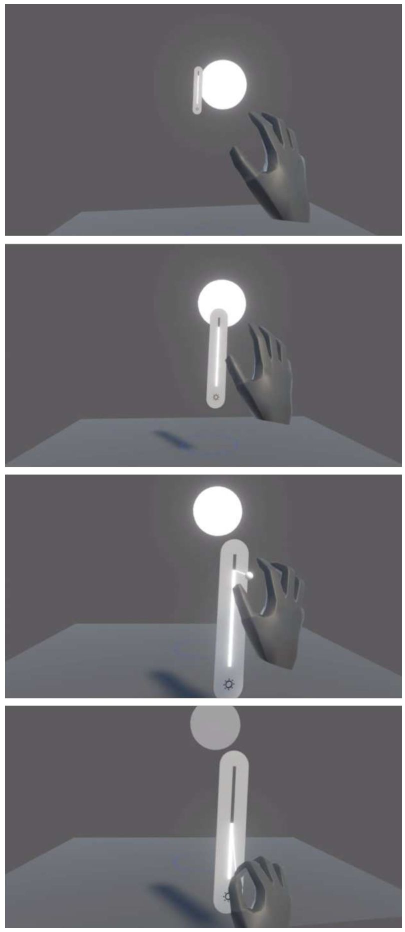


Figure 4.10: *transition an UI panel over distance for easy access*

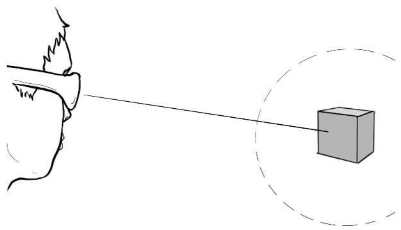


Figure 4.11: *when a user's eye gaze is close enough to an interactive object, assuming that a user is attempting to interact with it*

puts down his hand facing other orientations.

One of the challenges in this concept is to obtain eye gaze information precisely. Currently, the prototype does not include eye-tracking technology, so it only uses a raycasting from a user's head orientation to detect the eye gazing, which is not entirely accurate. One method to compensate with the inaccuracy is when a user's eye gaze is close enough to an interactive object, assuming that a user is attempting to interact with it. However, this problem might be resolved with the implementation of eye-tracking.

Additional, avoiding triggering the transition accidentally is essential. We could provide visual cues for a user when the system detects he is looking at an interactive object. For example, popping-up a small information UI with subtle animation to draw a user's attention is an effective approach.

Furthermore, this interaction might feel cumbersome if a user interacts with several devices close to each other, especially if multiple UI panels flying to a user together. Besides, this concept can offer a user to have a close-up examination of the detail information and complicated interactions.

Findings and Insights

Through the first round of prototyping and testing, I synthesized a few high-level insights that informed my later design process, and might also be generally applicable to other types of gestural interaction design.

- *Gestural interaction design should consider not only the intuitiveness of the gesture, but also the comfort level of the gesture, and try to avoid a gesture that is likely to cause fatigue over time.*

- *Proximity feedback (e.g., changing scale, color) are perceived as the analog to the hover state in GUI, and it can provide users with visual (sound) cues for next affordable actions.*

- *Encouraging user's fingers to tap on each other, or touch on a physical surface can provide additional haptic feedback, and make the interaction feel more concrete and solid.*

- *The lack of tactile feedbacks may cause some erratic motion, but adding a mediator (like a handlebar) could ease the tracking inaccuracy and hand jittering, making the interaction feel more natural and easy to control.*

- *Hand interaction is usually designed for arms reached close-range interaction, thus transitioning the interactive UI to arm-reached space in time can make the UI more accessible.*

5

CHAPTER

PROTOTYPING AND ITERATIONS
–
CONTEXTUALIZATION & APPLICATION

CONTEXTUALIZATION AND APPLICATION

The first round of interactions mainly concentrated on the basic interactions, while the second round attempts to engage with semantic contexts and puts the basic interactions into applications. This chapter will discuss more the opportunities and constraints of interacting with multiple devices, including a smart speaker, smart lights, computer displays, and tablets. The interactions are either developed from previous concepts or new design paradigms that address the problems differently.

Interaction 7
Grab & Place

Grabbing and placing a virtual object is intuitive, but unlike grabbing a real object, grabbing a virtual object does not offer any real feedback, e.g., force, haptic or tactile, so other elements from the previously proposed “five elements framework” can to be a handful to enrich the experience, such as using sound to compensate for tactile feedback.

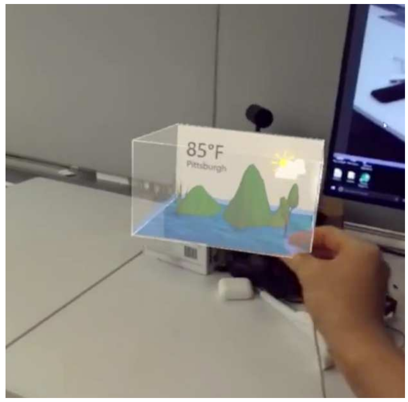


Figure 5.1: *initial design uses a bounding box for proximity feedback*

My initial design uses a bounding box for proximity feedback, but most users complain about its visual distraction, which sabotages the principle of invisibility discussed in the first chapter. As such, the later iteration pivots to a more subtle glowing outline, but still keeps the corner of the bounding box as the grabbing state, and it is proved to be quite effective for communicating the moveability of the grasped object.

Grabbing and placing can not only apply to virtual objects but also enable useful interactions for physical devices. One example is to place a virtual 3D music icon onto a smart home speaker to play a piece of music. I attached a predefined image marker on a CD cover, so the AR camera can recognize the cover, and displayed a related 3D icon. Placing a 3D music icon onto a smart speaker helps a user to build a conceptual relationship of how a physical object (a CD cover) turns into its digital representative (a 3D music icon), then it triggers a digital event (playing a piece of music) from a physical device (a smart speaker). Although it is very tech heavy underneath, the metaphor of grabbing a CD and placing into a CD player helps a user to contextualize it properly. Admittedly, many tested users do not discover the interaction initially, but it becomes self-explanatory after the first trial or a short instruction.

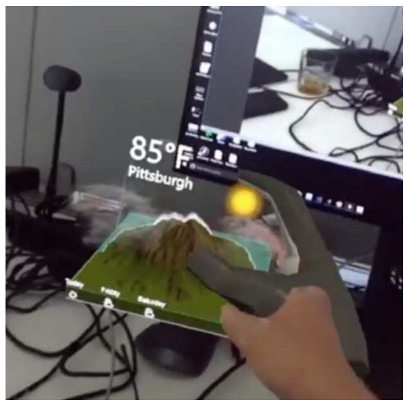


Figure 5.2: *a subtle glowing outline then replaces the bounding box to make it less distracting*

This approach takes a real-life physical experience as a metaphor and uses it digitally to inform an action that analogs its physical counterpart.

Figure 5.3: *mirror hand interaction allows a user to control a device over distance*

This paradigm may open up many new ways of how we interact with physical and digital information together in AR.

Interaction 8
Mirror hand

Opposed to the "transition UI panel" interaction discussed in the previous chapter, another way to interact with a device in a distance is to create a mirror hand at the target device. Very similarly trigger conditions, combining with the eye gaze and hand direction, we could first display an interactive UI next to the target device, and transition our mirror hand to it. Since the mirror hand moves and behaves accordingly to a user's actual hand simultaneously, it mostly feels like an extension of our hand and works very intuitively.

This concept can call back to the usage of a mouse. Although a mouse cursor and a user's hand moves in different space even in different axes, the control is still perceived to be responsive. That being said, the mirror hand does not work well if the target device is too far away, because it might be challenging to see the UI panel clearly or make precise depth judgment in such a long distance.

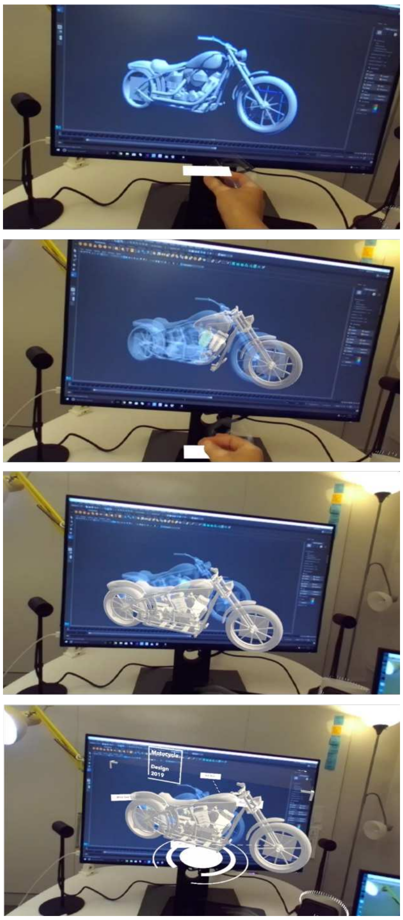


Figure 5.4: *sequence of pulling a 3D model out of screen to real-world environment*

Interaction 9
Pulling a model out of a computer display

Pulling a things out can metaphor a change of status. For instance, pulling a drawer out means an items previously hidden inside are now available for grabs. Pulling a door open allows us to access a room, and the status of a door changes as we pull it. So the interaction of pulling aligns with our existing mental models of switching status between two modes - inside/outside or open/close. Therefore, this concept utilizes this metaphor to afford switching modes between screen space (inside a display) and our physical environment(outside a display).

A user can pull out a 3D model from a computer display that he is currently working on, and place it in the real environment to view it three-dimensionally. Pulling action can communicate the mode changes between different devices, and it facilitates to ease the mental load of working with multiple systems.

Note that in this interaction, indicating the mode changes is critical. My approach is to change the shading and color of the 3D model inside the screen while a user is pulling it. By making the model inside transparent, the display conveys a meaning of loss or disabled, and it helps a user to build a mental connection of his action and the result. By harnessing a physical affordance of pulling and its metaphoric meaning, we could help a user to understand this unfamiliar way of switching mode from a computer display to AR space.

Interaction 10
Pulling sticky notes out of a tablet

Likewise pulling a model from a display, a similar approach can apply to switch modes between a tablet (e.g., an iPad) and the AR space as well.

Drawing or writing a sticky note in AR with gestures is difficult; thus a tablet can compensate for its limitations and create a more satisfying sketching experience. On the contrary, placing and arranging sticky notes for sense making, like affinity diagramming, on a 2D screen is quite limiting, but conducting it in 3D space is much more digestible and allows developing relationships across various pieces of information easily. Combining them seems an obvious choice, but the question is how to make the interaction and transition seamlessly?

In this interaction, a small handlebar sits aside of a tablet, and a user can grab and pull it out of the tablet, just like we pull a bookmark out of a book. Moreover, the sticky note inside the tablet also moves according to the pulling action. In this way, the two vastly different systems - a tablet and the AR space - are being connected seamlessly through a simple gesture. By harnessing the real world affordances that we are already familiar with, it built on our existing experience, as if we are interacting with a real-world object as opposed to digital information.

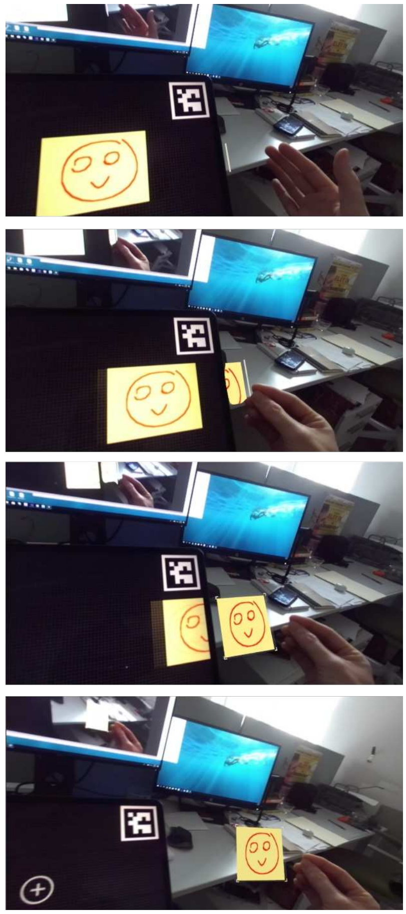


Figure 5.5: *sequence of pulling a sticky note out of an iPad to real-world environment for affinity diagramming*

CASE STUDY

Interaction 11
Slurping digital content with a virtual eyedropper

Unlike all interactions mentioned above, slurp takes a different approach that uses metaphor to create additional affordances for gestural interactions. This concept is inspired by the project "Slurp" (Zigelbaum, J., et al., 2008) from MIT Media Lab Tangible Media Group in 2007, and project "Slurp" creates a tangible eyedropper that allows communication of digital information between different devices. Similarly, this concept harnesses the metaphor of "slurping" and creates a virtual eyedropper

Figure 5.6: *slurp project from MIT Media Lab Tangible Media Group*

Figure 5.7: *use opening bottles as metaphors to trigger different pieces of music in project Tangible Bits*

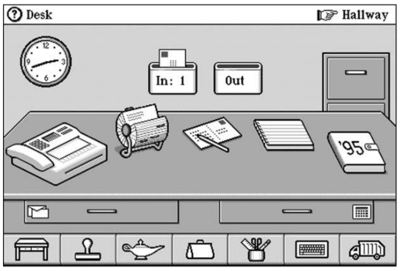


Figure 5.8: *desktop metaphor developed by Alan Kay at Xerox PARC in 1970s*

Figure 5.9: *slurping a 3D model*

with gestural controls to interact with digital content as well as physical devices, such as a smart light.

The metaphor approach is being used frequently in creating tangible interactions, especially from Hiroshi Ishii's project "Tangible Bits" (Ishii & Ullmer, 1997) and his followers. The fundamental question of metaphoric representation is the level of "perceptual coupling", as Ishii points out in "Tangible Bits: Beyond Pixels" (2008), and he suggests that "the success of a TUI often relies on a balance and strong perceptual coupling between the tangible and intangible representations." (p, xvii) The perceptual coupling can also apply to the AR interactions; the question of how close a metaphor to its abstract representation comes as the key to its success.

Moreover, metaphorical representation has long existed in the field of graphic user interface (GUI) design. In the 1970s, Alan Kay and his team developed the "desktop metaphor" at Xerox PARC, which considers the computer interface as a virtual desktop, imagining digital information being files, folders like their physical analogs sitting on our desks. The desktop metaphor often coupled with the paper paradigm, which imagines the digital interface as a piece of paper that can be marked on. The desktop metaphor and paper paradigm approach, as Key said, "describe a correspondence between what the users see on the screen and how they should think about what they are manipulating." Thomas Erickson, one of the essential brains behind Apple early Macintosh design, proposes that "metaphors serve as natural models; they allow us to take our knowledge of familiar objects and events and use it to give structure to abstract, less well-understood concepts." Moreover, this design philosophy still exists today in many different forms, such as skeuomorphism, Google Material Design.

Initially inspired by those projects and theories, I first created a virtual eyedropper that can slurp virtual objects, such as 3D models. The thought behind it is to extend the ability of an eyedropper from being able to slurp liquid, to slurping solid substances. The interaction allows a user to slurp and place objects in a distance. It is a big plus for gestural interaction, because gesture usually only deals with arm-reachable interactions. Besides 3D models, it can also slurp and place a virtual screen, and it could be useful if a user wants to set a video or browser to a place away from him or her.

Displaying an inventory of the slurped object locked to a user's hand can help memorize the content, and it also allows to slurp in multiple objects. However, slurping in various objects increases a user's cognitive load, and it is particularly frustrating if there is not an intuitive way for a user to select which to slurp out the next. Hence allowing slurping multiple objects adds a whole layer of complexity for a user to digest. Therefore, throughout later research, I mainly allow a user to slurp one piece of information at a time.

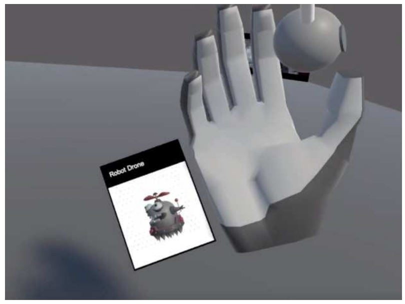


Figure 5.10: *display an inventory of slurped object locked to a user's hand*

Furthermore, a visual indicator is also necessary for providing clear visual cues of selected objects. A simple line is firstly introduced to visualize the pointing direction. Nevertheless, a similar problem of the pointing interaction (Chapter 3, interaction 1) occurs again that the inaccurate hand tracking and erratic motion causes the projected ray jittering. But since the slurping direction is calculated with the overall hand tracking, which is more stable than the finger joints tracking. In this case, a simple ease-in transition can be of a lot of help. Additionally, I added a calculated bezier curve to simulate a fishing pole snapping in a target object, which makes it feel a lot more responsive and intuitive. Lastly, since the slurp in and out takes the same gesture interaction - pinch, I employ a special shader effect with animation on the line to visualize moving in or out, but it turns out to be barely noticeable for most users.

Allowing the virtual eyedropper to interact with abstract information and smart devices can be even more magical. As an eyedropper is often used as a metaphor for color-picker in a lot of design tools (e.g., Photoshop), so it drove me to explore a new concept of picking up a color from our physical environment and apply it to a smart light. This abstract concept is also inspired by the deluminator in the Harry Potter movies. Employing an animated particle effect of color being slurped in and out, it helps to an embodiment this intangible process into a visualized substance, that is both intuitive and surprising.

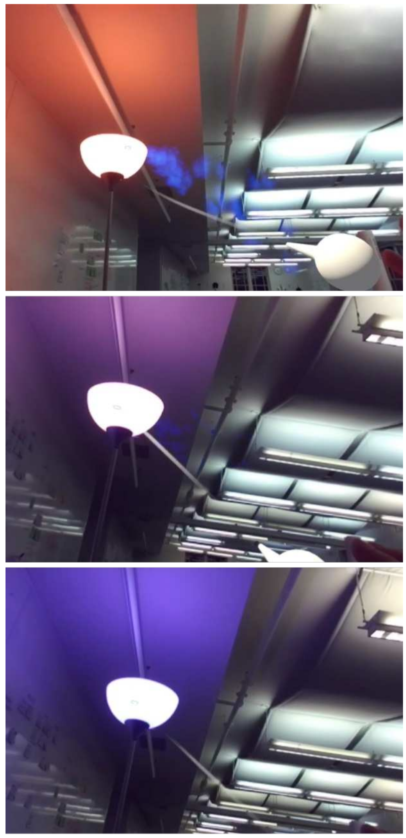


Figure 5.11: *pick up color from real-world, and slurp it to a smart light*

The metaphor expands the boundaries of what hand gestures can afford. It works as a mediator or as a tool that extends our hand capability. But designers need to be aware that users have different mental models of slurping an object versus an attribute of an object, and it means it would be confusing when both are enabled simultaneously.

All in all, for intangible, abstract or complex digital interactions, embodiment and materialization sometimes can be unexpectedly effective. It bridges the gap between the physical and digital by mediating the interactions. Although the metaphor does not always guarantee the intuitiveness for a user to understand without any learning in advance, it does indeed create a smoother learning curve that supports users to build from existing knowledge or experience that they are already familiar with.

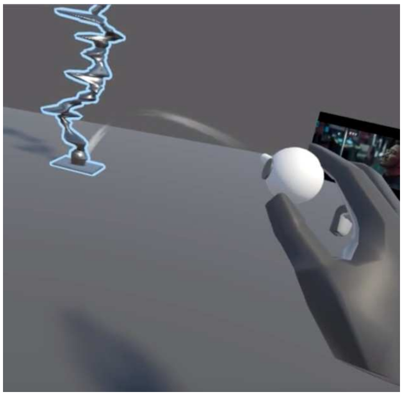


Figure 5.12: *drawing a curved line to mimic the physics of a fishing pole can help align users' mental models of AR interactions with real-world experience*

Key Insights:

By synthesizing the second round of prototyping and iterating, I summarize a few high-level insights that might generally be applicable for AR interaction design and future research.

• Mimicking real-world physics for virtual content can help align users' mental models of AR interactions with their real-world experiences, making it feel intuitive and satisfying.

• Besides direct gesture affordances, metaphors (e.g., a virtual eyedropper) can extend the affordances of gestural interactions, even creating magical interactions and encourage more playful exploration in an AR world.

• When using the slurp metaphor, designers need to be aware that people have different mental models of slurping an object itself versus an attribute of an object. It's confusing when both of them are enabled at the same time.

• Hand view allows manipulating attributes or adjusting details of an object; world view allows developing relationships across objects (e.g., grouping, organization, sense-making). We could leverage both to support different purposes.

• Physical affordances use users' existing mental models of interacting with objects, and it can help to ease their mental transition from one device to the other when interacting with multiple devices.

6

DESIGN RECOMMENDATIONS & DISCUSSIONS

CHAPTER

DESIGN RECOMMENDATIONS & DISCUSSIONS

For my research, I ultimately synthesized five high-level design recommendations that might be useful for later designers and researchers. Some of the recommendations are not only applicable for AR gestural interface design but can be helpful in general AR interface design. Besides this, I will also discuss some considerations and limitations of this research, which potentially could be future work to continue the study.

***Figure 6.1**: five-element framework of manipulating perceived affordances*

1. Consider the five elements framework for tweaking users' perceived affordances

My initial research demonstrates a five elements framework that affects a user's perceived affordances in AR, and they are "shape," "color and shading", "light", "haptic and acoustic" and "motion". Those five elements together shape how an object or a UI component is perceived.

Each element functions serve different purposes, but they are all interconnected. For example, light can be used for emphasizing or call for attention, while light, shading, and color could all communicate hierarchies, which sometimes act a similar purpose. When designing an interaction, designers need to consider them together. There might be many ways to convey a particular affordance, and it is difficult to value one to the others. With this framework, a designer can leverage it as a guide to start with, or examine whether there is some missing design consideration that can be utilized for his design.

2. Harness metaphors to create additional affordances in gestural interactions

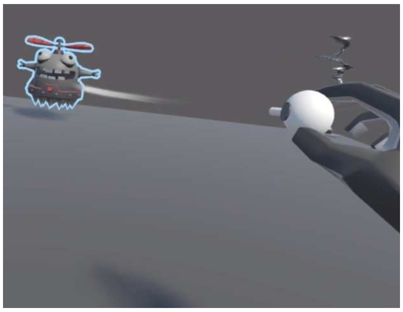


Figure 6.2: *harness an eyedropper as metaphor to extend gestural affordances*

Metaphors are compelling when they embody an abstract concept appropriately. Using metaphors with the hand can extend the affordances and user capabilities, and it creates many new opportunities for gestural interfaces. We can imagine those metaphors as tools that we use with our hands, such as a magnifying lens, a nail or a piece of paper. Think about what they can represent, metaphor, and affordances for interactions.

A useful exercise for finding a proper metaphor is first to understand the flow or purpose of the abstract system function that needs to be represented. Then spend some time brainstorming what objects, tools, natural phenomenons that are similar to it, and the things should be very familiar to your users. Then speed dating with your users to get early feedback before diving into it. As Ishii points out that the "perceptual coupling" of a metaphor and its representation is the most critical. Once a metaphor can adequately reflect the perceptual coupling, it helps to enlighten the abstruse into the obvious.

Managing users' expectation of a metaphor precisely in the first place can reduce users' confusion. A metaphor is an analog, rather than an accurate representation of a function, so the scale of what a metaphor can represent is usually vague. It requires us to set up users' expectation carefully. For example, in the slurp interaction, a virtual eyedropper function as a metaphor to slurp in/out information, but what kinds of information are slurpable is quite ambiguous. It, therefore, becomes very confusing when a user can choose both slurps in a virtual object and color from the environment. It is because a color is essentially an attribute of an object, and a user has different mental models of slurping in an object versus slurping in an attribute of it. But it turns out to be self-explanatory if we set the expectation clearly upfront that only enables to slurp in color. As such, setting up the boundary of a metaphor is necessary for intuitive interaction.

When designing metaphors for gestural interactions, we also need to be aware of the limits of gestures. Some metaphors might be proper as a representative, but it might be fatiguing to pose, or difficult to track accurately. Another technical constraint is that holding an AR object in hand sometimes does not look entirely realistic; especially we need to hold it inside the hand, it's hard to get the AR occlusion correctly. Those are the technical constraints that cannot be neglected by AR designers.

3. Switch modes between hand-space and world-space for seamless interactions

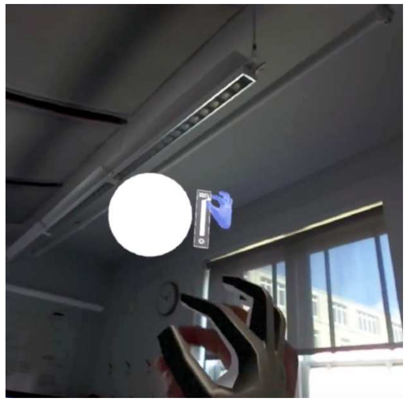


Figure 6.3: *use mirror hand interaction to switch between hand-space and world-space*

Hand-space and world-space represent two different paradigms of interactions. Each of them offers some values while also contains tradeoffs. For example, hand-space allows users to have a precise control and detail tweaking; whereas world-space can indicate the relationship of a virtual content to a real-world environment. In general, designers should consider to leverage them together for different purposes, in order to achieve a seamless experience in AR.

Hand-space and world-space firstly are targeting different distances of interaction. Hand-space is close-up interaction, and it refers to the interactions that either lock to a user's hand directly or always within a user's hand-reachable distance. If we consider users' ergonomics, digital objects in AR need to be very close to a user's hand, and a user should feel comfortable to interact with it. For example, a watch worn on a user's hand can be considered as in hand-space. Those are fixed to a user's hand appears like something he is wearing, or almost like existing as an extension of our body. On the contrary, world-space refers to content that is locked to our environment; it is usually far away from a user's hand and does not follow his hand movement. For instance, a clock on the wall could be treated as in world-space, and it exists as a part of the real environment.

Figure 6.4: *world-space allows a user to conduct affinity diagram in AR*

Hand-space and world-space present two different mental models for a user in AR space. Hand-space content is usually more accessible, and it follows a user, affords detailed close-up interaction. It typically draws more attention from a user since it is so close to a user. We can often put the content requires close-up detail tweaking, or quick access in a hand-space, such as a home menu or frequently used shortcut. On the other hand, world-space content stays in the distance, exists in our peripheral vision and is usually less noticeable. It builds a relationship with our environment, often presents meaning around it. For example, a button in the world-space close to the light is usually perceived as for controlling the device. However, we need to realize these two modes are also interchangeable. A tablet held in a user's hand is in hand-space, while once placed on a table, it becomes a world-space content.

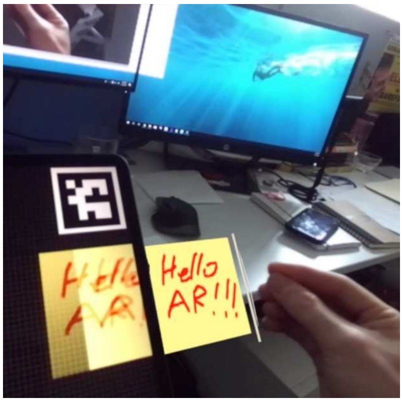


Figure 6.5: *Deluminator in Harry Potter movie uses animated particles to embody the light*

The interchangeable nature of these two modes allows us to switch modes for different interaction scenarios. For example, when a user intends to control a smart light in a world-space, we can bring the interactive UI to the user's hand-space allowing an accessible interaction. Another concept of mirror hand goes the other way around; it moves a user's hand to world-space to control the target light, which also creates a novel way of interacting far-distance device. Furthermore, in the interaction 5 (transition UI panel) mentioned above, a hand-space UI functions as a shortcut or quick access for a user, and it always follows a user's hand and becomes an extension of his body. Switching back to a world-space builds the relationship with the environment, and creates a spatial memory for a user to remember.

Additionally, switching modes between hand-space and world-space also serve for different purposes. A hand-space interaction allows manipulating attributes or adjusting details, while a world-space can be useful for developing relationships across objects, such as sense-making, grouping, and organization. In the tablet sticky notes example, a user sketches on a tablet in his hand-space, which enable high-fidelity drawing and writing; then he can pull the sticky note out to place it in world-space for conducting spatial affinity diagrams. This two-mode switch demonstrates how both of them can be useful for various purposes and could facilitate a seamless experience for users.

4. Leverage real-world physical affordances in AR to align users' mental models between different devices

People naturally assume the laws of physics, and affordances from the real world will also function the same way as we are in an AR environment. That's why we always see a user immediately touch and grab the virtual content after him putting on an AR headset. But a user's mental model of interacting with different devices may vary dramatically, and an effective strategy is to leverage the real-world affordances into AR world.

Applying our real-world physics to virtual content not only creates a more immersive and realistic experience, but helps signal specific actions that are afforded. In the slurping example, the visual indicator becomes exceptionally significant to signal when a user can slurp. In the initial design, I added a straight line to connect the center point of the slurped object to the virtual eyedropper in a user's hand, but the straight line doesn't communicate a sense of locking and snapping. I am inspired by the affordances of a fishing pole, where the top of the fishing pole will band with spring force to give a sense of weight. I simulated a similar physics force to the pointing line, and it felt much more reactive and satisfying.

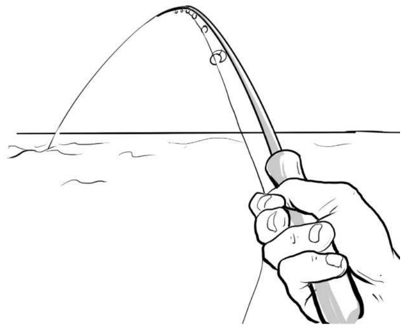


***Figure 6.6**: a fishing pole curves when*

Harnessing real-world physical affordance can align users' mental models among different devices as if the boundaries of them do not exist. A good example is pulling a 3D model from a screen. Because the action between those two devices(a computer display, and AR headset) aligns with each other, and it is already familiar in our physical interaction. So a user can seamlessly switch modes between a 2D screen, and a 3D AR environment. A similar approach occurs in the example of pulling a sticky note out of a tablet, and I employed a spring force to the sticky note. When a user is dragging it, the virtual sticky note inside the tablet moves along with the AR sticky note, and it bounces back when being released before fully dragged out, this interaction is as if a user is pulling a real note out of a book. It uses our familiar physical affordances to merge the boundaries of the two devices and creates a magical experience.

***Figure 6.7:** Deluminator in Harry Potter movie uses animated particles to embody the light*

Since people are all familiar with the physical affordances in our daily lives, it is easy to harness it to ease users' cognitive load of interacting with different devices. This strategy can be interrupted as an extension from the skeuomorphism philosophy from the early age of GUI design: rather than skeuomorph an icon from a physical object, it skeuomorphs the physical interaction into AR world. It is quite beneficial since the AR interface is still new and unfamiliar for most people.

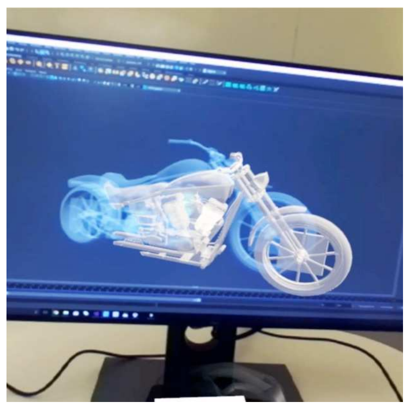


***Figure 6.8**: pulling a 3D model from screen to actual environment*

5. Empower the abstract with embodiment and materialization

Interacting with multiple devices can be conceptually abstract, but embodying and materializing the abstract process can help a user to understand it better. By bringing something invisible, abstract or non-exist into a physical or visual form, we could turn the unfamiliarity and the abstract into things we can perceive directly, or we know how it works.

Figure 6.9: *Deluminator in Harry Potter movie uses animated particles to embody the light*

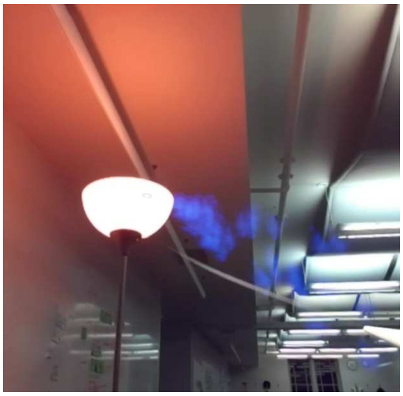


Figure 6.10: *The picked color is materialized in the final prototype*

The embodiment and materialization do not have to be real or accurate, but they need to reflect the essence of the abstract. It is beneficial for visualizing a complicated process and making information more digestible. In the example of slurping a color to a smart light, of course, there is a lot of technical detail behind the scene, but a user does not have to see it. A user only needs to understand color is picked up from where he or she is pointing at, then that color is slurping out directly to a smart light as intended. My initial design does not include any embodiment and materialization; the light changes immediately when a user slurps towards it. However, the process feels like a black box and takes a few times for a user to figure out what is essentially happening. For the later iteration, inspired by the "deluminator" in the *Harry Potter* movie, I added a little particle effect to materialize and embody the slurping process, and the particle inherits the slurped color, then flies towards the target light. The color of the light only changes when the color particle reaches it. With this simple materialization, it helps a user to understand the meaning behind it immediately, making it feel quite magical and satisfying.

We understand our physical world better because we can perceive it directly, and we usually apply what we can see, hear, or feel to interpret those abstract concepts, complex processes or overwhelming information. Embodiment and materialization are aimed to bridge the gap and help people decipher those abstruse into something we can digest quickly. In our real lives, there is so much information that is hidden, missing or non-exist, such as wi-fi signals, the amount of data in a USB drive, energy in a battery, or the data flow between different devices, which all can be better presented to users. Through embodiment and materialization in AR, we can see it, hear it, even interact with it directly, as if it is also a part of our physical world. It also helps to merge the boundary between virtual and physical, creating a seamless transition from what it is, to what we can interpret and interact.

limitations

This thesis research is not about designing for currently available technology, but for a future where augmented reality is ubiquitous and accessible for most people. Additionally, I assume that AR headsets will be lightweight and comfortable for people to wear a long time. We can already see signals in the current tech industry that spatial computing and ubiquitous computing are moving rapidly towards this future. Today, scannable QR codes are widely used to access the unique identity and information from a product, but in this future, every object would function as a “QR Code”, allowing people to interact with it through the lens of AR, as Kevin Kelly envisions the future that every physical object will “have its digital twin”, and interconnected through networks. This envisioned future lays a fundamental assumption of this thesis, and many interactions may reach its full potential when this future comes to reality.

Culture variations for gestures

The same gestures may vary in different cultures, but in this research, I attempt to avoid the influences of cultural variations in terms of the meanings of gestures. Instead, I look more at the gestures that employed from our physical interactions, such as grab, press, drag, pull, etc. Those gestures should be unaffected by cultural context and are generally applicable for larger audiences.

Limits of test users

This research focuses more on the exploration of technology assessment and prototyping instead of user-centered evaluation, and the goal is to explore and propose new possible design paradigms for AR gestural interactions. Therefore, current user tests are very preliminary, and future work should involve more participants and conduct more analytical evaluations.

Slurp data (files, photos)

In the slurp interaction, I prototyped working demos that can interact with virtual objects (3D models and virtual screens), colors from the real environment and smart lights. But I believe slurping with other types of data would also be valuable, for example, a file from a display or USB drive. In addition, since the AR headsets use cameras to capture the environment in real-time, it is possible to interact with physical photos, images, or documents as well, which would allow a user to manipulate digital and physical information more engagingly.

TERMINOLOGY

TERMINOLOGY:

Perceived affordances

Donald Norman's (1988) book The Psychology of Everyday Things introduced the term affordances into the vocabulary of design and human-computer interaction. Norman proposes that affordance is the design aspect of an object which suggests how the object should be used. Later, Norman (2008) clarified the term "affordances" as "perceived affordances," as it emphasizes the significance of people's perception rather object's actual affordances, and he added "What the designer cares about is whether the user perceives that some action is possible". (Norman, 2008, p. 3), so this thesis will mainly use Norman's definition of perceived affordances as reference for discussion.

Gestural interactions

A type of user interaction, where a user interacts with computer devices through a set of gestures. The gestural interactions in this thesis particularly refer to hand-gesture interactions, which can evolve the direction of hand, finger movement etc.

Head-mounted display (HMD)

A head-mounted display is a display device, worn on the head or as part of a helmet, that has a small display optic in front of users' eyes.

Mental models

In 1943, when Kenneth Craik (1943) first proposed the concept of mental model. He discussed a relationship of our internal representation and external world. Craik argues that the"internal model of reality — this working model — enable us to predict events which have not yet occurred in the physical world" (p.82). He suggests that the process firstly translates the external information into words, numbers, or other symbols, and then through the process of reasoning, deduction and inference, the symbols are subsequently retranslated as a bridge to help people understand the external world. (p.82) Later, the most widely accepted notion of a mental model is "internal representations of system in a particular knowledge domain. These internal representations are formed through knowledge (instruction) or experience or a combination of the two". (Staggers, 1993, p. 601) I used this definition for this thesis.

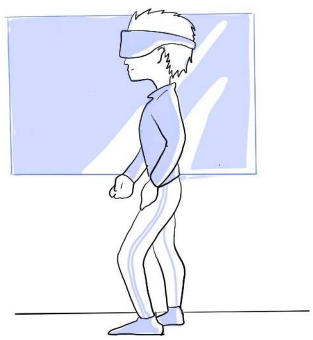


Figure 7.1: *body-locked - digital content is fixed to a user's environment*

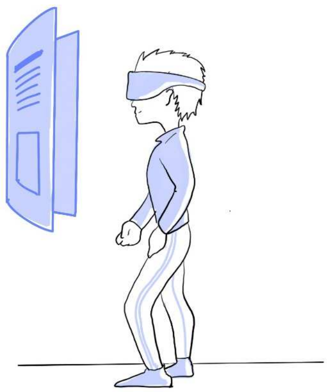


Figure 7.2: *body-locked - digital content moves along with a user's body*

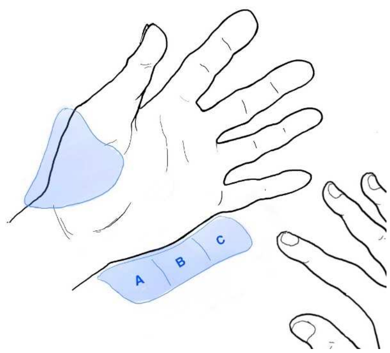


Figure 7.3: *hand-locked -digital content is fixed to a user's hand*

World-locked,body-locked and hand-locked

The concept of world-locked, body-locked and hand-locked are used to describe the different modes of displaying information in augmented reality and virtual reality. It distinguishes various ways of showing virtual information relative to the user's present position and real-world physical space.

World-locked

When the virtual information is static relative to a user's physical surroundings, it is typically considered to be world-locked, which means the virtual content is locked to the real world space. In this case, the information will keep static until the user intentionally moves it. More broadly speaking, we can consider anything that is placed in our physical environment to be world-locked.

Body-locked

By understanding the concept of world-locked, we could then apply the same way of description to the idea of body-locked, and body-locked means the information or content are static relative to a user's body; therefore it can be thought as locked to the user's body. However, since a user's body also moves from part to part, so typically the virtual content is fixed to either user's head, or the center of a user's body, which can be both regarded as body-locked.

Hand-locked

Likewise, hand-locked means the virtual content or information are static relative to a user's hand, for example, a ring we are wearing, or a watch on our wrist can both be treated as hand-locked. Note that the ring is locked explicitly to the finger which we wear the ring on, but the watch is on our wrist, and they are not precisely the same motion space, but they both, in general, can be defined as hand-locked or we can call it in hand space.

BIBLIOGRAPHY

BIBLIOGRAPHY

Caudell, T. P., & Mizell, D. W. (1992, January). Augmented reality: An application of heads-up display technology to manual manufacturing processes. In Proceedings of the twenty-fifth Hawaii international conference on system sciences (Vol. 2, pp. 659-669). IEEE.

Craik, K. J. W. (1943). The nature of explanation. Oxford, England: University Press, Macmillan.

Ens, B., Quigley, A. J., Yeo, H. S., Irani, P., Piumsomboon, T., & Billinghurst, M. (2017). Exploring mixed-scale gesture interaction. SA'17 SIGGRAPH Asia 2017 Posters.

Eugene Krivoruchko and Keiichi Matsuda, Designing Cat Explorer (2018), http://blog.leapmotion.com/designing-cat-explorer/

Gaver, W. W. (1991, April). Technology affordances. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 79-84). ACM.

Gibson, J. J. (1979). The ecological approach to visual perceptions.
Ishii, H., & Ullmer, B. (1997, March). Tangible bits: towards seamless interfaces between people, bits and atoms. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (pp. 234-241). ACM.

Ishii, H. (2008, February). Tangible bits: beyond pixels. In Proceedings of the 2nd international conference on Tangible and embedded interaction (pp. xv-xxv). ACM.

Leap Motion (2016). Ergonomics in VR Design, http://blog.leapmotion.com/ergonomics-vr-design/
Lo, D., Lockton, D., & Rohrbach, S. (2018, April). Experiential Augmentation: Uncovering The Meaning of Qualitative Visualizations when Applied to Augmented Objects. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 490). ACM.

McGrenere, J., & Ho, W. (2000, May). Affordances: Clarifying and evolving a concept. In Graphics interface (Vol. 2000, pp. 179-186).

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE TRANSACTIONS on Information and Systems, 77(12), 1321-1329.

Nancy Staggers & A. F. Norcio (1993). Mental models: concepts for human-computer interaction research. Washington DC: International journal of man-machine studies

Norman, D. A. (1988). The psychology of everyday things(Vol. 5). New York: Basic books.

Norman, D. A (2008). Affordances and Design, https://jnd.org/affordances_and_design/

Radcliffe, D., Watson, E., Grint, R., Rowling, J. K., & Warner Home Video (Firm). (2011). Harry Potter and the Deathly Hallows: Part 1. United States: Warner Home Video.

Spielberg, S., Molen, G. R., Curtis, B., Parkes, W. F., Bont, J. ., Frank, S., Cohen, J., ... Twentieth Century-Fox Film Corporation. (2003). Minority report.

Stereolabs, https://www.stereolabs.com

Zigelbaum, J., Kumpf, A., Vazquez, A., & Ishii, H. (2008, April). Slurp: tangibility spatiality and an eyedropper. In CHI'08 Extended Abstracts on Human Factors in Computing Systems(pp. 2565-2574). ACM.