Beyond the Desktop

Metaphoric Interaction Alternatives for Emerging AR Technology

Ashley Cook

Department of Graphic and Industrial Design
College of Design, North Carolina State University

May 3, 2022

Submitted in partial fulfillment for
the degree of Master of Graphic Design

Denise Gonzales Crisp, Committee Chair
Professor of Graphic Design, Director of Graduate Programs

Tasheka Arceneaux-Sutton, Committee Member
Associate Professor of Graphic Design

Jarrett Fuller, Committee Member
Assistant Professor of Graphic Design

ABSTRACT

As tech companies continue to invest in Augmented Reality Smart Glasses (ARSG), 2D screen interface design practices will not remain sustainable for 3D field-of-view interfaces. Previous introductions of new interaction systems, such as the desktop computer, implemented concrete metaphors to represent functionality. This project explores how utilizing the user's prior knowledge of sensorimotor and cultural experiences can allow for more flexible metaphors. Specifically, the studies investigate possible gestural and oral interaction metaphors, as well as functionality metaphors, for user tasks in order to inform principles for a future ARSG system.

ACKNOWLEDGMENTS

THANK YOU,

To my incredible committee: Denise, Tasheka, and Jarrett

To all of the encouraging NCSU College of Design Faculty

To the most supportive people I know: Ashley, Carl, and Andrea

To my voice actors and best friends: Andrew, Sarah, and Ashley

To my family and undergraduate professors who believed in me

TABLE OF CONTENTS

Introduction

> *"What we call common sense—the body of widely accepted truths—is, just as Heidegger and Nabokov thought, a collection of dead metaphors. Truths are the skeletons which remain after the capacity to arouse the senses—to cause tingles—has been rubbed off by familiarity and long usage."* (Rorty, 1989, p. 152)

As humans, we cannot get away from metaphors. Metaphors are intrinsic to how we communicate, how we think and perceive, and in turn, who we are (Lakoff, 1980). Design, as a form of communication, is then metaphorical. Since the advent of the personal computer, researchers have related the computer interface to paper metaphors. Paper as a metaphor for Graphical User Interfaces (GUIs) has been around since Doug Engelbart introduced the paper paradigm in his research on *Augmenting Human Intellect* in the 1960s (Engelbart, 1962). It is an easy connection for users to make as both paper and screen-based interfaces are two-dimensional and relatively the same size. Interestingly, Xerox, a paper-based company, created the desktop which is a metaphorical paper-based system, as a part of the Xerox Alto personal computer in 1973 (Thacker et al., 1979). The screen of the Alto took the paper metaphor quite literally by making it the same size as a vertical US letter sized paper. The paper paradigm also imports many typographic and hierarchic print standards into the interface—headings are (usually) placed at the top, margins inhabit the sides, etc. The cursor could even be seen as a digital form of the manicule. Although computer interfaces have abstracted a bit since the Xerox Alto, even Google's Material Design uses material as "the metaphor" to "reimagine the mediums of paper and ink" (Material Design, n.d., Principles section).

Heading into the near future, emerging technologies such as Augmented Reality (AR) and Virtual Reality (VR) give designers the opportunity to expand further into the three-dimensional world. Not many AR/VR design practices exist that do not rely on the precedents of a two-dimensional interface (Rauschnabel, 2018). Spatiality brings many metaphoric implications into interface and interaction design ripe to be explored by designers, which are investigated in the following studies.

1.1. PROBLEM STATEMENT

The effectiveness of metaphors in interface design is a contentious topic. Even those who favor metaphors agree that there are limits to its efficacy (Sease, 2008). However, not all metaphors are created equally. Several studies have categorized human-computer interaction (HCI) metaphors in terms of strengths and limitations. The examples discussed in the papers "Elastic metaphors" and "Computers as people" are concrete, functionality, interface, and interaction metaphors (Khoury et al., 2004; Fineman, 2004). Concrete metaphors use everyday objects familiar to the user as the target domain, defined as the *thing* or the *vehicle* used to understand a more abstract source concept (Khoury et al., 2004; Lakoff, 1980). Because the source and target domains are so closely tied together with concrete metaphors, its concreteness inherently limits the interface design to a set of rules specific to the target domain (Blackwell, 2006). If the digital object breaks away from what the physical object can do, then users can become confused. The desktop metaphor, composed of its synecdochal *files*, *folders*, *wallpaper*, and *trash can,* is a concrete metaphor. The trash can famously breaks the metaphor by allowing disks to eject —something a physical trash can cannot do.

These concrete metaphors also limit the user by inadvertently defining who the user should be. For example, in the desktop metaphor, the user is a worker. The limitation also extends to what a user should be doing with the interface. The designer is limiting user choice and simultaneously placing the user into a role that they may not occupy. Current users are not only doing work-related productive tasks on their desktop now, but are also fulfilling roles such as a shopper, artist or gamer, as well as doing tasks uninvented at the time of the design of the original desktop like browsing the web.

The taxonomy used in this investigation divides HCI metaphors by its type of relationship to the interface, categorizing them as: functionality, interface, and interaction metaphors (Fineman, 2004). Functionality metaphors "deal with what the product can do" while interface metaphors "deal with the mechanics of how tasks are accomplished…[and] are *expressions* of the underlying metaphor" (ibid, pp. 6-8). For example, the *desktop* is the functionality metaphor, while the icons of *files* and *folders* are the interface metaphors that are expressive of the desktop. As discussed above, functionality and interface metaphors can be problematic if created as concrete metaphors. Interaction metaphors, in contrast, relate to the "interaction between human and computer" and are "task independent" (ibid, pp. 8) For the desktop, direct manipulation within a WIMP (windows, icons, menus, pointer) system is the interaction metaphor as it relates data to physical objects that can be interacted with. The user is an agent in control of manipulating the passive computer. While WIMP metaphors (such as *selecting* a function from a *menu*, or *scrolling* through a *page*), or direct manipulation metaphors (such as *drag and drop* or *point and click),* may be restrictive, the interaction system within which it is built combines with its input forms—a mouse/trackpad and keyboard—and compounds its inefficiency (MacKay, 2003). Keyboards were purposefully designed to be inefficient back when the keys could become jammed and the design has not changed (Alden, 1972). Alternative text inputs are also an area ripe to be explored as voice input becomes more popular (Kiseleva et al., 2016). Other WIMP system pitfalls include: the time-consuming process of navigating through a menu to select a command; the overreliance on icons to conceal information, forcing users to learn its meaning; and the window unnecessarily acting as an enclosed boundary within an already constrained screen (Nielsen, 1993; Gentner et al. 1996).

Touch devices, while not WIMP devices, are still tied to the screen's boundaries. Touch is an effort to move from traditional WIMP interfaces, but the designs still rely heavily on user familiarity with WIMPs and tend to favor similar features (Jetter et al, 2014; Blackler, 2015). Many smartphones, such as the iPhone, are built around the "home screen" metaphor. Users navigate different paths away from the home screen to interact with an ecosystem of applications. Apple's home screen metaphor is strengthened through built-in app interface metaphors like the clock, phone, contacts, notes, and calendar. The iconic representations for these apps are depictions of older technology, such as an analogue clock, landline phone, contact book, lined notebook, or paper calendar. Their design is a continuation of the desktop design approach of using dated concepts to facilitate understanding within an abstract space. Soon, the smartphone iconography (and desktop iconography such as folders) will be as unrecognizable as the floppy disc, which continues to represent saving files.

As users become more familiar with digital tools, concrete interface metaphors may no longer be needed. The desktop metaphor was, at least partially, influenced by Jean Piaget's and Jerome Bruner's psychological and pedagogical theories from the 1970s (Blackwell, 2006). The theories helped the researchers at Xerox PARC justify a metaphorical design as an intuitive introduction of the personal computer to novices (ibid). The reasoning behind its implementation is no longer the reality as current generations are interacting comfortably with computers and post-WIMP devices such as smartphones and tablets. Furthermore, the concept of a novice / expert divide needing different interfaces is founded on shaky ground (Raskin, 2000). Since intuitive systems are learned systems, the user must learn how to interact with the system whether it is metaphorical or abstract, simple or complex (Hurtienne, 2007). I'm not advocating for abstraction and complexity over metaphor and simplicity, but for alternatives to the systems in place now.

1.2. JUSTIFICATION

Previous work on alternatives for the desktop have explored spatiality, integrating physical realism, and imitating molecules into the digital world, as well as different types of input devices such as pens, touch, and gestures (Agarawala, 2006; Lee, 2013). However, these experiences are still reliant on the screen or desktop computer as the primary device while partially incorporating 3D or AR elements into the design (ibid). The evolution of a portable device that allows for a variety of tasks to be accomplished can be found in emerging AR technology (Shin, 2021; Rauschnabel, 2018). AR smart glasses (ARSG), in particular, have many advantages: 1) the device is a lightweight, wearable technology that can be carried anywhere; 2) digital elements enhance the surrounding environment instead of detracting users away in the way that a current computer interface does; 3) digital elements can be placed anywhere within the environment within a reasonable scale and can be transported to other locations; 4) the device engages the users' entire range of view and can allow for peripheral engagement as well; 5) it allows for more embodied, immediate, and immersive interaction to occur (Banky et al, 2017). As this new technology develops, the first batch of interaction systems and interface metaphors will be loosely based on the current interface paradigm because it is familiar to the user. My investigation, instead looks at ARSG functionality and interaction metaphors that might form once that paradigm is mostly dropped. With a shift to ARSG, the use of physical keyboards might diminish as a multimedia-based culture continues to shift to an age of "secondary orality" where electronics allow oral communication to be deliberately chosen over literary modes of communication (Ong, 1982, pp. 3). The development of voice recording, multimedia sharing, and other electronic oral technology has allowed and will continue to allow a shift from text-based input to verbal input. The implications position ARSG to become a much more embodied, oral, and flexible mode of interaction for the next generation of users.

The limitations discussed for the WIMP system and desktop metaphor such as an overreliance on metaphoric iconography and concrete metaphors can be avoided within new frameworks for AR interface design. Researchers are studying how interfaces can be intuitive while still remaining novel through the use of conceptual metaphor or image schematic design (Hurtienne, 2015). Image schemas are the recurring patterns of our physical and bodily experiences in the world, such as up–down or near–far. Intuitive design is based on a user's prior knowledge, which is why technology that has been drenched in society is intuitive. Hurtienne suggests that prior knowledge coming from cultural or sensorimotor experiences can be used as the backbone for interface design instead of literal, physical objects (2007). Using cultural or sensorimotor knowledge might prompt metaphors that are much more flexible than concrete interface metaphors.

1.3. ASSUMPTIONS AND LIMITATIONS

The studies documented here are placed within a plausible, speculative near-future scenario. This future avoids the paradigms and design practices of screen-based interface interactions that are prevalent now. The experiences, although probably not intuitive to the contemporary user, are intended to be easily learnable for future experienced users (Earnshaw, 2018). Every new system that is introduced does have a learning curve, and these studies are no exception.

For the purposes of this project, emphasis is placed on gestural and oral interactions, and the functionality metaphors of the interface. Interface metaphors and the accompanying visual design is not the focus of the studies, but designing some graphical interfaces was necessary to express interactions, but were not the focus of the project. With the given time constraint, an entire AR-lens interaction system cannot be created. Instead, I am creating limited scenarios to produce principles that could inform a future system. The scenarios propose alternative modes—not single solutions—of interaction and input compared to WIMP and touch devices today. By using a design-based discovery model, the studies aim to engage designers in a critical dialogue through variable explorations (Peterson, 2021). The designs will propose *what could be* in a future reliant on ARSG and the implications of that reality in contrast to the contemporary desktop model.

The studies imply that the technology needed for the advanced interactions will be created by this future time. The technology includes AR capabilities, but also refers to machine learning, computer processing capabilities, bone conduction, computer vision and others. Some instances could require additional wearable devices to work that will not be shown, such as wearable hand trackers.

Each study focuses on a different type of user—the term *user* is used loosely here as the future will most likely have a new way of defining the type of engagement happening between person and machine. A universal system that works for every user is unattainable, but I am assuming that the users are technologically competent and able to interact through gesture and speech. As with the desktop metaphor, the design and interactions provided are limited to a Western cultural perspective. Other cultures will need their own culturally significant interaction metaphors for a system similar to this. Additionally, I acknowledge privacy and data concerns from systems that pull behavioral data, but this facet is not the focus for this project.

Background to the Problem Space

2.1. ANNOTATED BIBLIOGRAPHY

Five categories of select literature from the literature search are listed and explained in the table below (Table 2.1.1). ***Metaphors in HCI*** (Human-Computer Interaction) discusses the effectiveness of metaphoric interface design. ***Image Schemas and Intuitive Interfaces*** examines the definition of intuitive interaction and how image schematic design could create intuitive interfaces. ***Post-WIMP Interaction Systems*** explores alternative ways of thinking about and measuring the success of post-WIMP interaction styles. ***Theory Papers*** encompasses the main influential theories supporting this paper's framework. ***Toward the Future*** discusses speculative design futures and where technology and culture might be headed.

TOPIC	TITLE, AUTHOR, DATE
1 **Metaphors in HCI** Metaphor is used in digital interface design to bridge the gaps in users' knowledge when interacting with abstract digital interfaces (Blackwell, 2006). There are multiple taxonomies for categorizing metaphors within HCI, such as Fineman's which is used in this investigation (Fineman, 2004). Sease discusses in detail various other frameworks such as operational and organizational metaphors, also known as "noun" and "verb" metaphors (Sease, 2008). The desktop is one of the most recognizable HCI metaphors, and perhaps the most literal. Don Gentner and Jakob Nielsen proposed a new way of thinking about interface design by creating the antithesis of the Mac. The Anti-Mac is built on the reality of computers creating new paradigms no longer tied to outdated technology or physical re-creations (1996).	**Metaphor's Role in the Information Behavior of Humans Interacting with Computers.** Sease, R. (2008) **The Reification of Metaphor as a Design Tool.** Blackwell, A. (2006) **The Anti-Mac Interface.** Gentner, D. & Nielsen, J. (1996) **Elastic Metaphors: Expanding The Philosophy Of Interface Design.** Khoury, G. R., & Simoff, S. J. (2004) **Computers As People: Human Interaction Metaphors In Human Computer Interaction.** Fineman, B. (2004).
2 **Image Schemas and Intuitive Interfaces** Interfaces are not inherently intuitive, but instead is a learned interaction (Raskin, 2000). Interactions and user	**Towards Intuitive Interaction Theory.** Blackler, A. (2015) **Design for Intuitive Use - Testing Image Schema Theory for User**

experiences are defined as intuitive if the design is familiar to users. The familiarity comes from the user's past experiences with similar interactions (Blackler, 2015). This definition creates a paradox between what is intuitive and what is innovative: since intuition is learned and familiar, it cannot be new and innovative (Hurtienne, 2015). Hurtienne conducted a study to sidestep the paradox by arguing the use of image schemas in interface design could create an interface that is both intuitive and innovative (ibid). While Hurtienne's concept is founded on conceptual metaphor and image schema theories, Jetter's framework for "Blended Interaction" is based on conceptual blending theory (2014). Conceptual blending takes multiple input spaces and blends them together to create meaning (ibid).

Interface Design. Hurtienne, J. & Blessing, L. (2007)

The Humane Interface. Jef Raskin. (2000)

Blended Interaction: Understanding Natural Human–Computer Interaction In Post-Wimp Interactive Spaces. Jetter, HC. et al. (2014)

Designing with Image Schemas: Resolving the Tension Between Innovation, Inclusion and Intuitive Use. Hurtienne, J. et al. (2015).

3
Post-WIMP Interaction Systems

WIMP (windows, menus, icons, pointer) interfaces are a type of interaction system most famously used in desktops and laptops. Jef Raskin details how WIMP interfaces can have a lack of transparency through its heavy use of icons, and hidden structures of menus and windows (2000). Researchers have created frameworks to move beyond the WIMP paradigm for decades (Nielsen, 1993). However, instead of having new, different styles of interaction, many post-WIMP devices do not stray very far from the fundamentals of a WIMP interaction and are considered augmentations of it (Earnshaw, 2018). Although AR and mixed reality (XR) are older technologies becoming more popular, user experience research in the field is still limited (Shin, 2021). Much of the research on AR and ARSG has focused on education for students or for gaming (ibid; Banky et al., 2017). ARSG, as a novel device, has the potential to become the next big thing following the PC and smartphone (Rauschnabel, 2018). However, its lack of precedence for design and functionality requires more research (ibid).

Research and Development on Interfaces of the Future. In Research and Development in Digital Media. Earnshaw, R. (2018)

Virtually Enhancing The Real World With Holograms: An Exploration Of Expected Gratifications Of Using Augmented Reality Smart Glasses. Rauschnabel, P. A. (2018)

Does Augmented Reality Augment User Affordance? The Effect Of Technological Characteristics On Game Behaviour. Shin, D. (2021).

Investigating Affordances Provided By A Head-Mounted Augmented Reality Immersive Device For The Real-Time Online Supervision Of Experimental Learning. Banky, G. P. and Blicblau, A. S. (2017)

Noncommand User Interfaces. Nielsen, J. (1993)

4
Theory Papers

Embodied cognition views the mind as understanding reality from the body's interactions with the world

Affordance, Conventions, and Design. Norman, D. (1999).

The Ecological Approach to Visual

(Wilson, 2002). These repeating sensorimotor interactions within the world create image schemas, such as up-down or left-right, that are the basis for languages underpinnings (Johnson, 1987). Language is composed of conceptual metaphors that map abstract domains onto concrete domains (Lakoff, 1980). Speakers usually do not consciously realize these mappings. Tangentially related is Neisser's Spatial Perception Theory which views perception as a cycle of anticipating schemata from environmental information that continually shapes our actions (Neisser, 1978). The theory of affordances states that objects have a range of actions that a person can perform with the object (Gibson, 1979). In terms of design, perceived affordances are created to guide users into performing tasks considered necessary by the designer (Norman, 1999).	**Perception.** Gibson, J. (1979) **The Body in the Mind.** Johnson, M. (1987). **Metaphors We Live By**. Lakoff and Johnson. (1980). **Six views of embodied cognition.** Wilson, M. (2002). **Perceiving, Anticipating, and Imagining.** Neisser, U. (2005) **Cognition and Reality.** Neisser, U. (1976)
5 **Toward the Future** Walter Ong comprehensively discusses the evolution of oral societies to the literary societies of today, and the wave of "secondary orality" overtaking our literary society as electronic media allows for oral sharing (Ong, 1982, pp. 3). Design strategies for looking toward the future include Speculative Design practices, Design Fiction, and Ethnographic Experiential Futures, among others. Speculative Design looks at possible and plausible futures and its purpose is to open debate into what is wanted for the future (Dunne et al., 2013). Experiential Futures visually and experientially renders potential futures, and can be combined with Ethnographic Futures Research for a more in-depth study (Candy et al., 2019).	**Orality And Literacy: The Technologizing Of The Word.** Ong, W. J. (1982) **A Conceptual Uses & Gratification Framework On The Use Of Augmented Reality Smart Glasses.** Rauschnabel, P. A. (2018) **Turning Foresight Inside Out: An Introduction To Ethnographic Experiential Futures.** Candy, S., & Kornet, K. (2019) **Speculative Everything: Design, Fiction, And Social Dreaming.** Dunne, A., & Raby, F. (2013).

Table 2.1.1 **Annotated Bibliography**

2.2. PRECEDENTS

To inform the investigation, I researched and compiled a variety of existing precedents surrounding metaphor or mimicry within interface design, as well as alternatives to desktop and WIMP experiences. The precedents are categorized as Alternative Desktop Experiences, Common Metaphoric Design Systems, and AR and XR Interaction Systems.

ALTERNATIVE DESKTOP EXPERIENCES

BumpTop

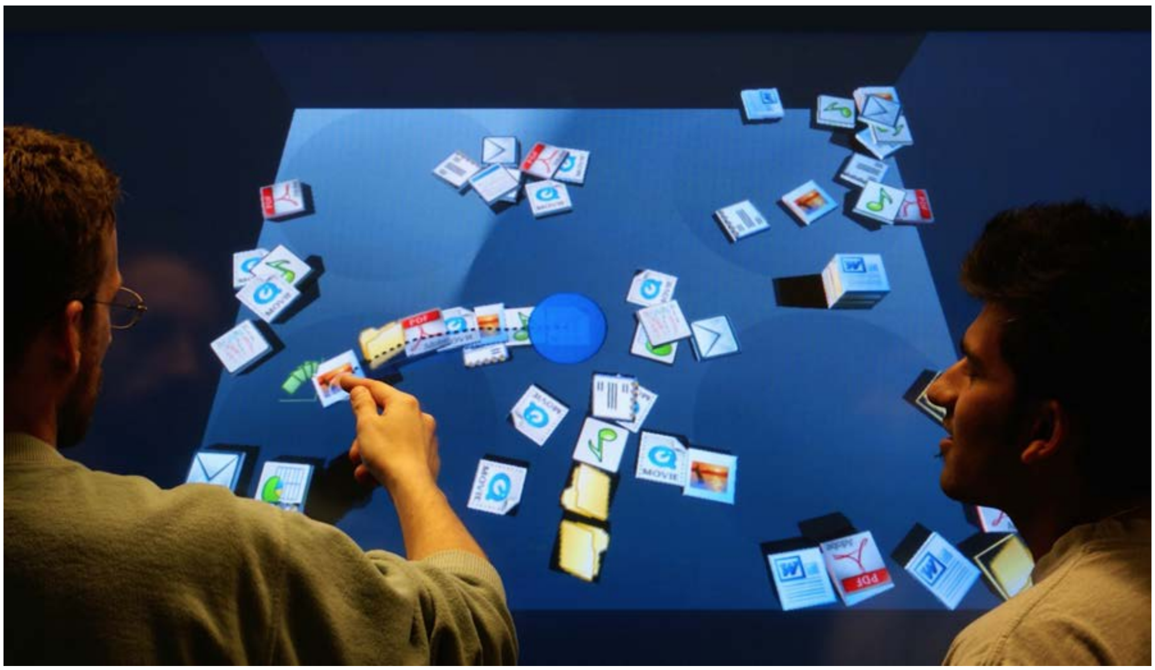


Figure 2.2.1 Screenshot from a video of BumpTop which mimics 3D physics and touch capabilities (Agarawala, 2016).

Anand Agarawala created BumpTop in 2006 for his master's thesis (Figure 2.2.1). BumpTop took the premise of the desktop metaphor, adding physics and three-dimensionality. Files are organized into low-effort stacks instead of high-effort filing folders as categorization (Agarawala, 2006). Objects can be manipulated to ram into one another, pinned to the virtual wall, and lassoed together. The system uses touch and a stylus, moving interaction closer to the post-WIMP realm, but still utilizes the keyboard. Bumptop changed the flat, 2D screen interface into a faux 3D environment, restructuring the desktop's abilities through spatiality and dimensionality.

SpaceTop

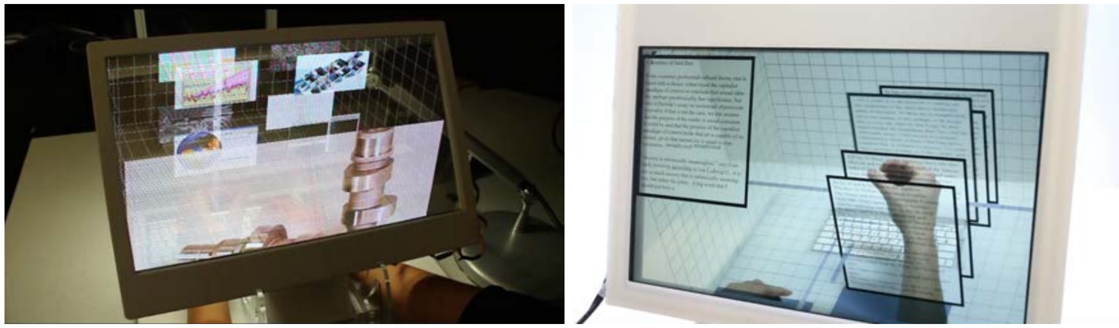


Figure 2.2.2 Screenshots showing SpaceTop's 3D translucent desktop screen (Lee, 2015).

Created by Jinha Lee after completing his master's degree at MIT, SpaceTop fuses 2D and 3D design to create an alternative solution to GUIs' limited screen space and makes the screen a more embodied experience in the process (Figure, 2.2.2). The interaction style shifts from a traditional pointer or touchscreen to a gestural interaction system behind a translucent screen. The gestural system favors hand input rather than that of a finger, similar to touch devices, and relies on the keyboard for text input. Interestingly, the design is taking the concept of ARSG technology and placing it into the context of a desktop computer screen as a workstation alternative. SpaceTop uses a sliding door metaphor to guide the user fluidly between the 2D productive space and the 3D organizational space (Lee, 2013).

Mercury OS

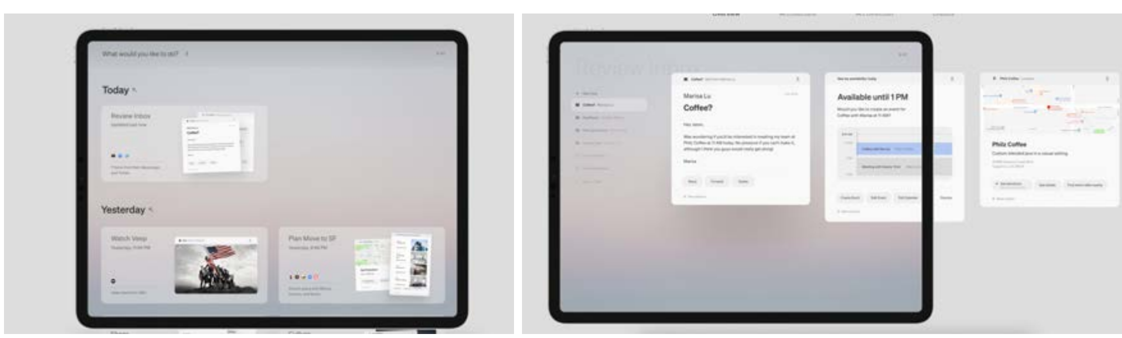


Figure 2.2.3 Mercury OS project modules (left) and view of a selected module (right).

Jason Yuan designed Mercury OS as a speculative operating system based on Jef Raskin's interface principles (Figure 2.2.3). The interface is described as humane, focused, fluid, and familiar. The system "rejects the Desktop Metaphor and App Ecosystem as inhumane," and does not support apps or folders (Yaun, 2019, Humane section). Mercury encourages information discovery through buildable modules that respond to user intent. Yuan designed the system to be intuitive so that users might easily engage with familiar elements. Mercury could be categorized as a near-future transitional interface as it is easily accessible to average users today. It utilizes touchscreen and voice interaction with an optional keyboard attachment.

COMMON METAPHORIC DESIGN SYSTEMS

Google's Material Design

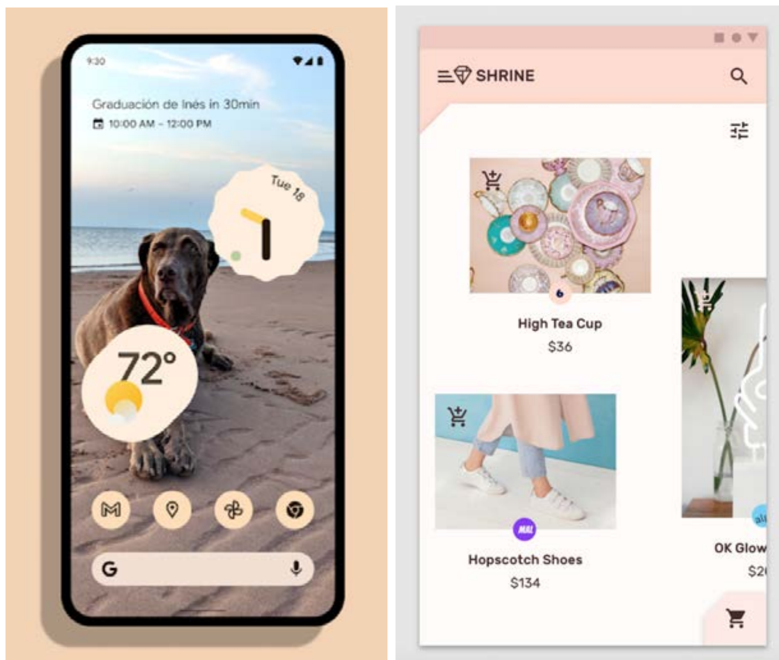


Figure 2.2.4 Screenshots of Material's interface examples (Material Design, n.d.).

Material is a design language for Android devices created by Google (Figure 2.2.4). The system describes material as "the metaphor" by "reimagining the mediums of paper and ink" (Material Design, n.d., Principles section). It tries to recreate the same effects users see in the physical world— light, shadow, and texture, for instance—within the digital world. However, the system is still very two-dimensional. The interface relies heavily upon icons, and adding companion text was not an option until the last update in 2021. Before then, the user was forced to learn the meaning of each icon which reinforces opaque, habitual interactions instead of allowing visibility into the interface elements.

Apple’s Skeuomorphism

Figure 2.2.5 Apple’s skeuomorphic iPhone design (left) versus the newer, flattened icon design (right).

Apple’s iPhone design system was once skeuomorphic, meaning the virtual objects mimicked the design of physical objects in a relatively realistic way (Figure 2.2.5). For example, the “Newsstand” mimicked an actual wooden bookshelf. Flat design eventually replaced the skeuomorphic style, with many designers referring to it as the end of the skeuomorphic style. However, most objects still mimic physical objects, now as more simplistic icons. The underlying system still retains desktop and home screen metaphors such as the wallpaper, clock, and notepad. There is an opportunity for applications to depart from physical reality metaphors and embrace more abstraction afforded by digital spaces.

The Cloud

The Cloud metaphor is understood by users without the need for reinforcement through visual interface elements. The Cloud is understood as an ether of data floating above users. Users can pull down, or download, files from the Cloud, or upload files into the Cloud. Beyond the up-down schema and mental model of a floating data bin, the Cloud does not explain what it is or does. The only visual presence of it within the interface is the cloud icon. The metaphor shields users from understanding how cloud computing works, as most other metaphors in HCI similarly do. The hidden functionality is not detrimental to its ability to be used and begins to set a framework for flexible metaphors that can still be based on physical objects.

AR AND XR INTERACTION MODELS

Microsoft HoloLens

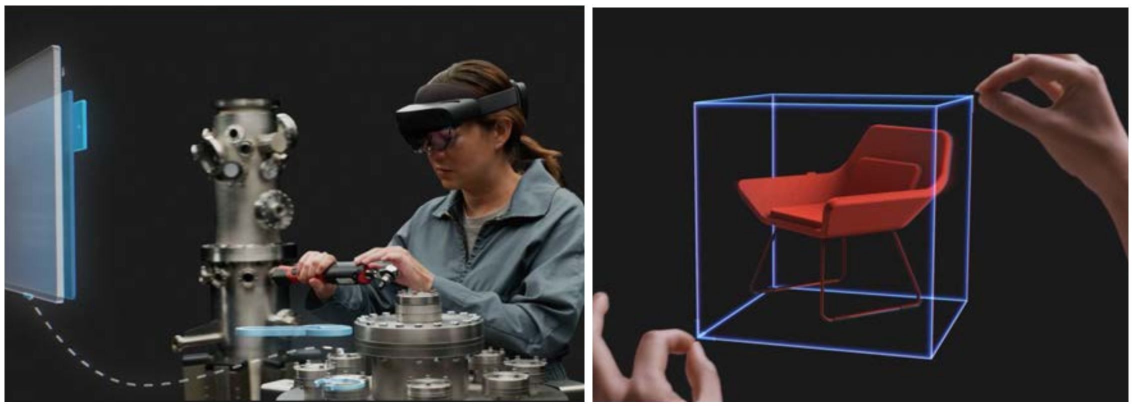


Figure 2.2.6 Examples of HoloLens in a manufacturing setting (left) and a gestural interaction example of scaling an object (right) (Microsoft, n.d.).

HoloLens is an ARSG, although Microsoft refers to it as mixed reality or holograms rather than AR (Figure 2.2.6). Microsoft released the HoloLens in 2016 for developer use and has since released the second generation for commercial use, although it has not been released to average consumers as of this writing. HoloLens projects 2D interfaces and 3D educational elements into the environment, making it a marketable transitional technology for institutions. Microsoft markets HoloLens towards three specific markets: education, healthcare, and manufacturing. The industry limitations are most likely due to the limited codebase in the field.

Lenovo ThinkReality A3

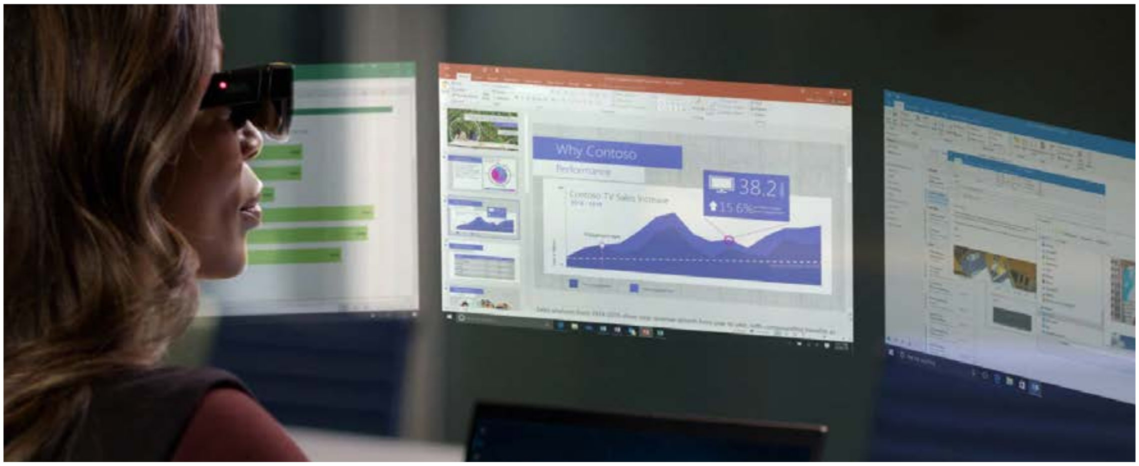


Figure 2.2.7 Mockup of ThinkReality A3 glasses from the users view (Lenovo, n.d.).

Lenovo ThinkReality A3 glasses target office workers and are an extension of the computer screen (Figure 2.2.7). Instead of having an additional monitor, users can see multiple augmented windows within their field of view. The glasses also work in limited manufacturing settings. The user interacts with the mouse and keyboard from the computer and controls the movement of the augmented screens through head gestures. The system is simple and persists with the desk worker paradigm.

Ultraleap Paint

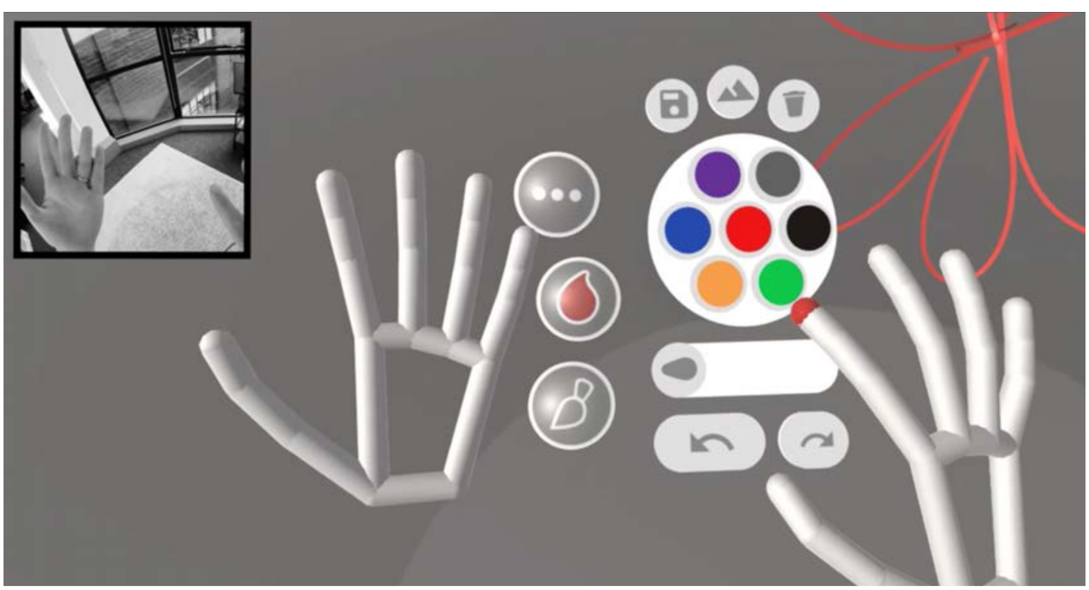


Figure 2.2.8 Demonstration of Ultraleap's Paint VR experience (Leap Motion, n.d.).

Ultraleap uses handtracking to create hands-free gestural interactions in XR / VR experiences. The Paint application is a virtual paint studio that requires a VR headset plugged into a PC (Figure 2.2.8). Since the environment is entirely virtual, the hands are recreated, and the background is free of distractions. In contrast, AR does not afford a distraction-free background which poses problems for visual contrast between the interface and the background. The interface is simple and still relies heavily on iconography and buttons derived from current visual interface languages. To open the paint interface, for example, users face their left palm up as if they are holding a physical paint palette. The right forefinger grabs a dab of paint from a limited selection of colors, adds it to a mixing pile, and then can add more dabs of different swatches to mix the paint. The interaction mimics real-life paint mixing and does not use a traditional digital color wheel.

KEY INSIGHTS FROM THE PRECEDENTS:

Paper and ink metaphors do not necessarily have to steer the interface into leaning heavily on 2D design.	Systems requiring overly organized structures may inadvertently cause disorganization or hinder low-effort users.
Typography within XR environments remains flat, most likely due to limited exploration into alternatives for legible, spatial type.	Support of app ecosystems or filing systems is not necessary for creating the next operating system.
Post-WIMP systems in the current market are using the WIMP system as a precedent to that which users find familiar and intuitive in order to create transitional interfaces.	Iconography for newer media that mimics older technology places the icon system into a specific user timeframe that can quickly become outdated.
Interface metaphors do not have to be visual, and can instead create mental models for users to understand.	Interfaces function as the middle man between the user and complex system underneath. The translation between the two entities will always limit user capabilities.
Icons can conceal information and, in some instances, information that might be better suited as words.	XR gestural interactions are mainly based on the learned knowledge of WIMP systems / buttons or sometimes mimic cultural / physical experiences.

Investigation Plan

3.1. CONCEPTUAL FRAMEWORK

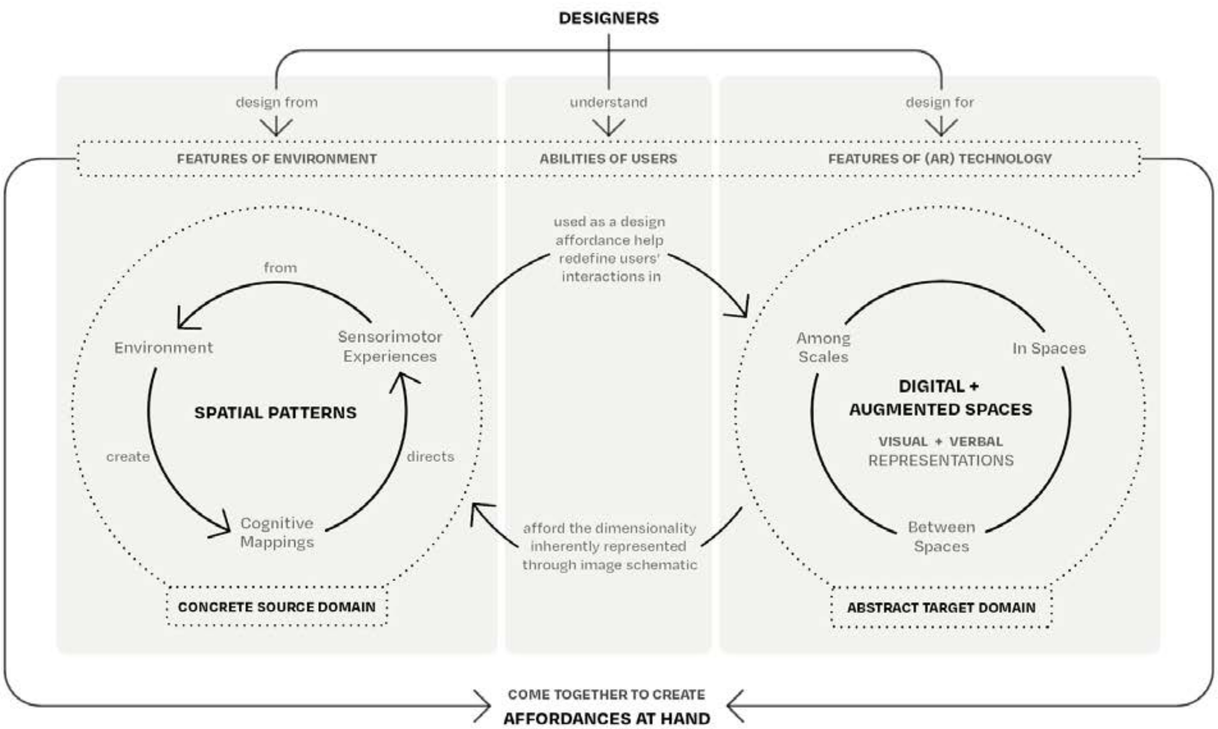


Figure 3.1.1 Mapping Spatial Affordances of Augmented Environments Framework.

The conceptual framework (Figure 3.1.1) underpinning this study is based on: the Theory of Affordances (Gibson, 1979; Norman, 1988); Conceptual Metaphor (Lakoff and Johnson, 1980), and Image Schema Theory (Johnson, 1987); and the Perceptual Cycle Model (Neisser, 1976). In order to go beyond the stylistic use and functional replication of the WIMP paradigm within an AR environment, precedents specific to AR need to be created for suitable features and functionality. Designers can explore more novel AR features and functionality by mapping spatial patterns, sensorimotor experiences, and cultural knowledge—mostly generalizable to people of the same culture—onto augmented interactions and interfaces.

THEORY OF AFFORDANCES

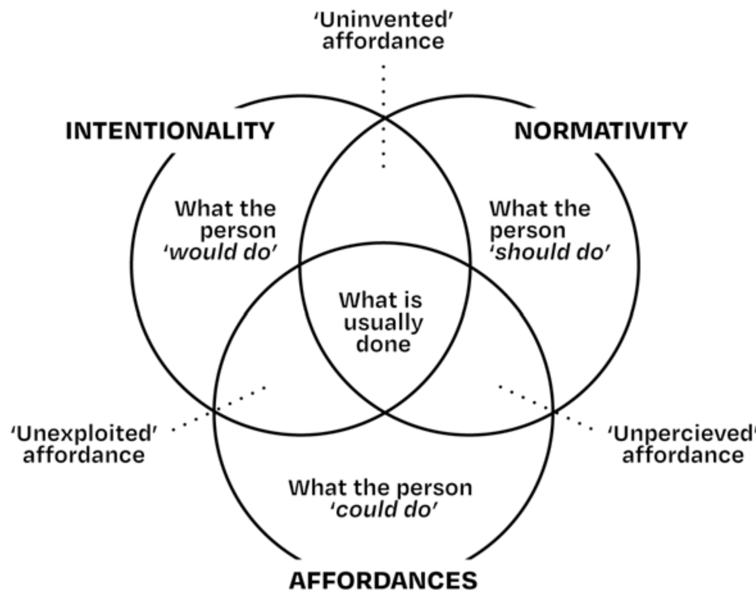


Figure 3.1.2 A sociocultural model for an affordance theory of creativity (Glavenau, 2012, p. 197).

Redrawn by Ashley Cook

James Gibson, a prominent American psychologist of visual perception, devised the Theory of Affordances. The theory was later adopted by Don Norman, an influential design researcher and professor, and retrofitted into a design perspective to explain how users interact with artefacts. Gibson argues that perception is not only shaped by the form of an object or spatial relationships, but also by an object's affordances (Gibson, 1979). He defines affordances as all the possible actions that can be taken with an object (ibid). Norman furthers the theory with *perceived* affordances, defined as actions that "the user perceives...[as] possible" (Norman, 2004). Not all affordances are perceived, but all affordances are technically possible. Norman gives the example of clicking a graphic object on an interface: the screen affords clicking anywhere, but does the user perceive that clicking on the object is "a meaningful, useful action to perform?" (ibid, pp. 1). The theory has since been built upon to create the Affordance Theory of Creativity, a framework incorporating sociocultural influences that might limit the number of perceived affordances of an object (Glăveanu,2012).

CONCEPTUAL METAPHOR AND IMAGE SCHEMA THEORY

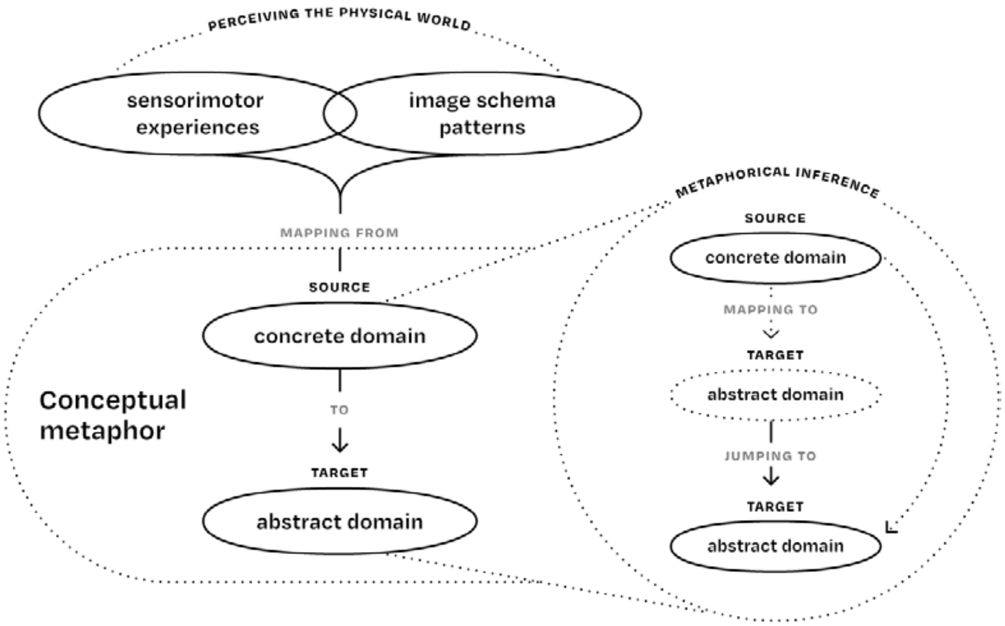


Figure 3.1.3 Conceptual Metaphor and Image Schema Model.

Drawn by Ashley Cook

Conceptual Metaphor and Image Schema Theory are based on the premise of Embodied Cognition. According to Embodied Cognition Theory, a person's conceptual cognitive system is grounded by their perceptions of physical experiences and body movements (Wilson, 2002). The recurring sensorimotor patterns of those physical and bodily experiences create image schemas, such as front-back or up-down, that are used as the basis for conceptual metaphors (Hurtienne et al., 2007). Conceptual metaphors are central to humans' thought processes to help explain everyday realities. Abstract domains are mapped onto concrete domains derived from the physical environment (Kövecses, 2020). For example, the conceptual metaphor GOOD AS UP metaphorically maps GOOD, an abstract idea, onto UP, a concretely understood idea, in the phrase "we hit a peak last year" (Lakoff et al., 1980, p.16).

Center—Periphery	Up—Down	Front—Back
Near—Far	Left—Right	Scale
Location	Path	Containers

Table 3.1.1 Examples of Image Schematic Spatial Patterns.

PERCEPTUAL CYCLE MODEL

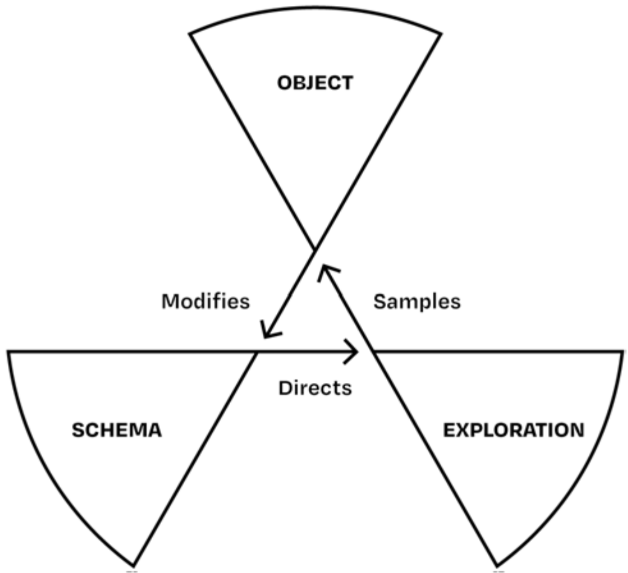


Figure 3.1.4 The Perceptual Cycle (Neisser, U. 1976).

Redrawn by Ashley Cook

Ulric Neisser, a notable psychologist, focused on perception and memory, conceived the Perceptual Cycle Model (Figure 3.1.4). The model states that top-down processing and bottom-up processing rely on and revolve around each other (Neisser, 1979). Perception is a cycle of anticipating schemata and environmental information that in turn, shapes a person's actions (Neisser, 1978). Thus, sensory experiences are neither completely internal nor external, and perceiving is active rather than passive.

3.2. RESEARCH QUESTIONS

RESEARCH QUESTION:

How can the design of metaphoric gestural and oral interactions within a speculative ARSG interface enhance task-appropriate activities for various everyday users through the technology's affordances in spatial patterning?

SUBQUESTIONS:

1. ***Productivity.*** How can the gestural and oral interactions of productivity-driven, text-based tasks be redesigned within an AR environment in a future that favors oral input, collaboration and access to physical, and digital information?

2. ***Creativity.*** How can interactions with visual and oral creative tasks allow for a transparent process of dimensionally making within an AR interface that allows for augmented collaboration and communication between viewer and artist?
3. ***Communication.*** How can gestural and oral interactions within an AR interface allow for fluid interpersonal communication, both synchronously and asynchronously, with others?

DEFINITION OF TERMS

Accessive tasks	Tasks allowing the retrieval and dissemination of data or information with a machine.
Affordances	Means by which actions are possible within an interactive or static object that establish a relationship between the object and the user.
Contextual tasks	Tasks relevant and specific to the activity at hand.
Communicative tasks	Tasks relating to the exchange of information or ideas between users.
Collaborative tasks	Participatory or cooperative tasks between users working together.
Gestural interactions	Body movements used to initiate actions in the interface.
Interpersonal communication	Expressional social communication between two people.
Metaphoric interactions	Interactions understood in terms of specific user knowledge from something else.
Oral interactions	Vocal / verbal sounds or non-lingual noises used to initiate actions in the interface.
Productivity-driven tasks	Tasks for efficient production of information output; traditionally within an office software suite.
Spatial patterning	Image schematic patterns (such as front-back, left-right, near-far) formed from bodily interactions and experiences in the physical world.
Speculative interface	An interface that could exist within a possible future scenario.
Transparent	Not withholding or hiding information in the interface.

Task-appropriate activities	Relevant and purposeful interface activities within the context of the target user's computer usage and the individual user's contextual needs.

3.3. INVESTIGATION MODEL

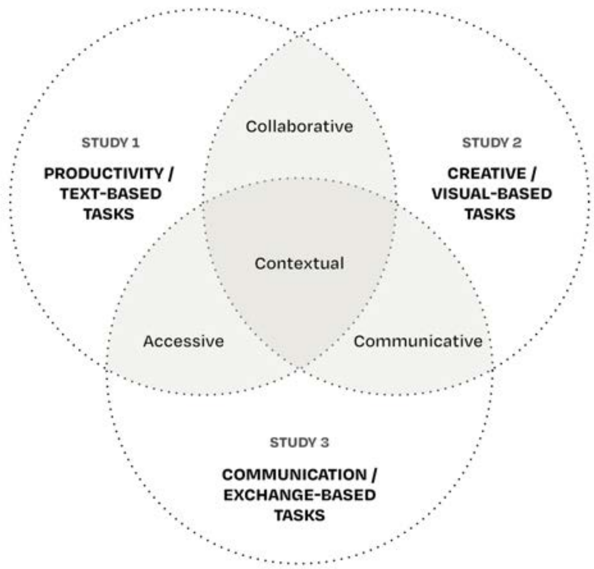


Table 3.3.1 Investigation Model each referencing a study-based subquestion.

My investigation model is divided into three studies associated with three subquestions (Table 3.3.1). All three studies explore contextual activities within a scenario specific to a persona with two additional themes (accessive, collaborative, or communicative activities) explored. The chosen tasks surrounding each study emphasize normalized digital activities in the hopes of informing future studies on how a "home" interface could function in an ARSG environment.

3.4. SCENARIO

For each study I wrote a scenario for a specific persona. The section regarding each study will go into more detail on the corresponding scenario coupled with the persona.

PLAUSIBLE FUTURE SCENARIO

By the 2020s, personal computers had splintered off into numerous smart devices. The frenzy innocently began with the advent of a personal desktop computer decades before and eventually branched off into laptops, tablets, smartphones, smartwatches, and other responsive devices. Laptops and smartphones were beginning to share information through bluetooth and Cloud computing. However, a user's personal data was still somewhat segregated between devices.

The spectrum and jumble of personal mini-computers and device hopping are now obsolete and mostly replaced with a single portable device: AR smart glasses (ARSG). A user's personal information cloud now follows them through the glasses, which means that no matter where the user is, whether or not another device is in use, their digital data is always available.

This transition has changed how a user interacts with their data through the glasses. Users need not lug around a handheld device. Dictation and handwritten input are more popular than keyboard input, and most children are no longer required to learn to use the keyboard. Listening has trumped long-form reading as well. The shift in culture from mostly literary screen-based media to oral media is ensured. The preference to listen, see, and speak has made digital repositories such as the web—rooted in written documentation–obsolete.

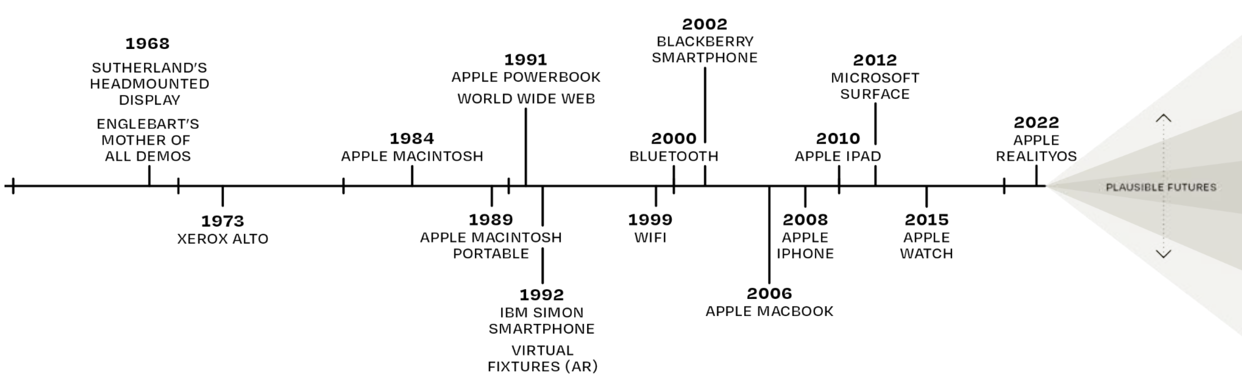


Figure 3.4.1 Plausible Future Timeline.

PHASE 1 **Screen-based technology**	**PHASE 2** **Transitioning Technology**	**PHASE 3** **Augmented Reality Technology**
AR complements the screens already widely available	AR primarily acts by itself in a variety of screen types	AR has its own conventions based on its specific affordances
AR is additive to the interface	Acts complementary to traditional interfaces *and* acts separately from traditional interfaces	No more carryover of non-AR screen-based interface practices
AR is mostly viewed through a handheld device	AR interface design is still largely based on the user's familiarity with the screen	Multiple devices allow a variety of XR media interfaces

Table 3.4.1 Phases of transition from today's technology to a future saturated with AR technology. Each of the studies sits in between Phase 2 and Phase 3.

ARSG TECHNOLOGY

My investigations utilize AR smart glasses technology with bone conduction speakers placed at the ear. Bone conduction funnels sound through the skull bone into the inner ear, allowing the ear canal to remain open and receive sound. When the machine vocalizes words to the user, the sound manifests through bone conduction speakers. In addition, hand tracking capabilities and spatial computing power register the hand's location in relation to virtual objects. In my plausible future scenario, these technologies have progressed to the point that spatially-aware ARSG interaction is possible, economical to use, and sought out by the general population.

Studies

4.1. Productivity-Driven, Text-Based Tasks

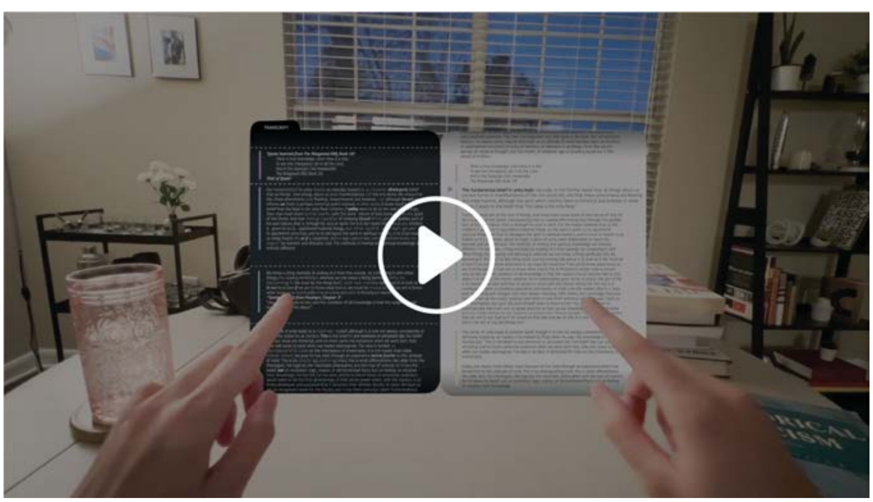
The first study takes traditional desktop tasks, places them into a near-future ARSG environment, and redesigns its functionality around the implications of the technology's affordances and the change in societal communication styles.

STUDY 1 PERSONA AND SCENARIO:

Olive is a 23-year-old grad student in the near future. She received her first smartphone at the age of 12 and was introduced to various AR apps. A few years later, companies began launching AR smart glasses that augmented the screens that consumers already used. Olive pre-ordered hers right away. Now a graduate student, she has ditched all her devices except for her AR glasses with a built-in microphone and bone conduction speakers. Olive primarily uses her device to complete schoolwork. The work consists of writing papers, which entails extensive research and editing, sometimes with remote collaborators. Writing has taken on a different form since the introduction of keyboard-less technologies. Many people, including Olive, prefer the process of dictation over typing with a virtual keyboard.

PROTOTYPED VIDEO

The video is linked below or can be viewed on YouTube at: https://youtu.be/mQxOftPRxEs



ALTERNATIVE LAYOUT MODES

The paper paradigm is heavily entrenched in current desktop productivity applications and works as a functionality metaphor or the overarching metaphor theme of the interface. Even the term "writing a paper" suggests that the final work will exist on physical paper. For example, if a user is typing in a word processor today, most likely, the design is WYSIWYG (What You See Is What You Get), meaning the form matches the output. The digital page replicates the printed version of the page. This study challenges the assumption that word processors should rely on print output or specific page count in the future and offers alternative typographic layouts.

Figure 4.1.1 shows a layout using traditional letter-sized papers arranged spatially around the user's workspace. Peripheral pages blur out of focus until the user gestures them towards the center of attention. The user swipes her full hand to recenter the pages, which makes use of a simple left-right schema, reiterated by the paper structure (Figure 4.1.2). For small type to be legible within AR, it usually must be placed onto a background. This limitation visually references the paper paradigm resembling typeset words on paper. In the iterations that followed, I kept the backgrounds for legibility purposes.

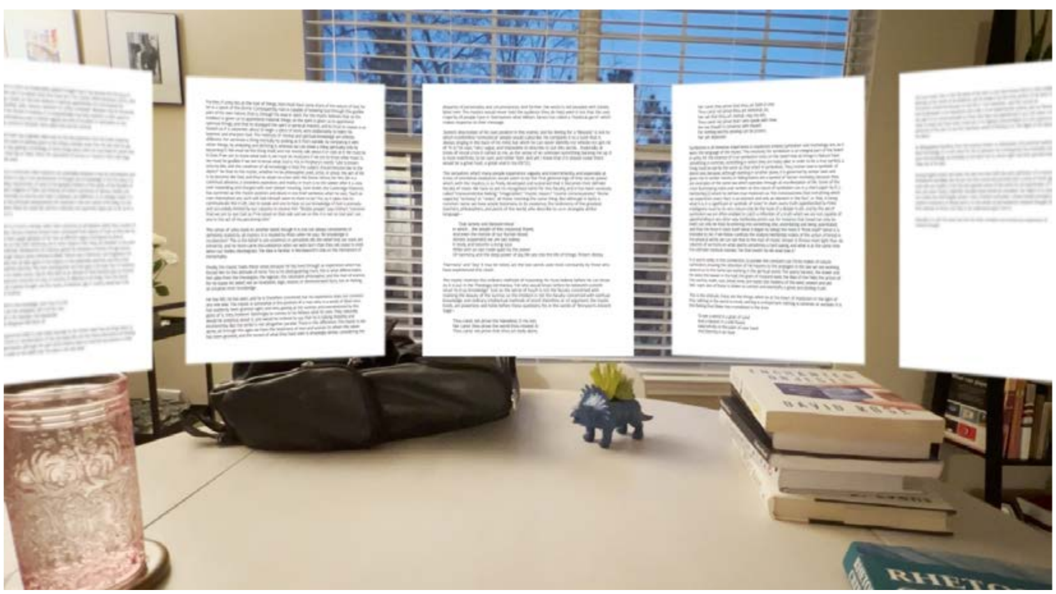


Figure 4.1.1 Traditional letter-sized paper arranged in a semicircle in front of the users field of view.

Figure 4.1.2 Full hand swipe from right to left pushes the rightmost papers towards the center of the frame (Left); Both hands brought together merges the paper into a scrollable view (Right).

To switch from a spread of papers to a condensed scrolling view, the user brings her hands together (Figure 4.1.3). This motion mimics separate pieces of paper meshing together into one and brings the left and the right sheets together towards the center. The gesture merges the papers into a scrollable view. Instead of having to zoom into an interface or increase the font size to get larger type, she can spread her thumb and forefinger diagonally apart to increase the size of the type without resizing the background perimeter (Figure 4.1.4). If she wanted to change the background's perimeter, her forefingers could take two opposing corners and readjust its size (Figure 4.1.4). She can spread her thumb and forefinger vertically apart to increase the line height (Figure 4.1.5). These gestures quickly allow for customization of the reading or viewing experience with little need to consider the output form.

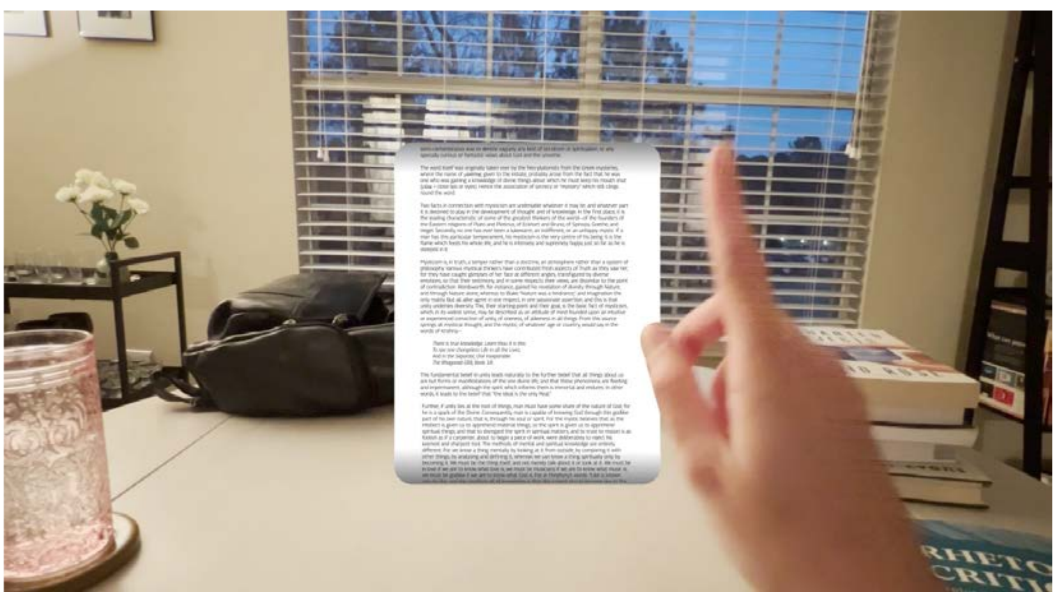


Figure 4.1.3 Merged pages in a scroll view.

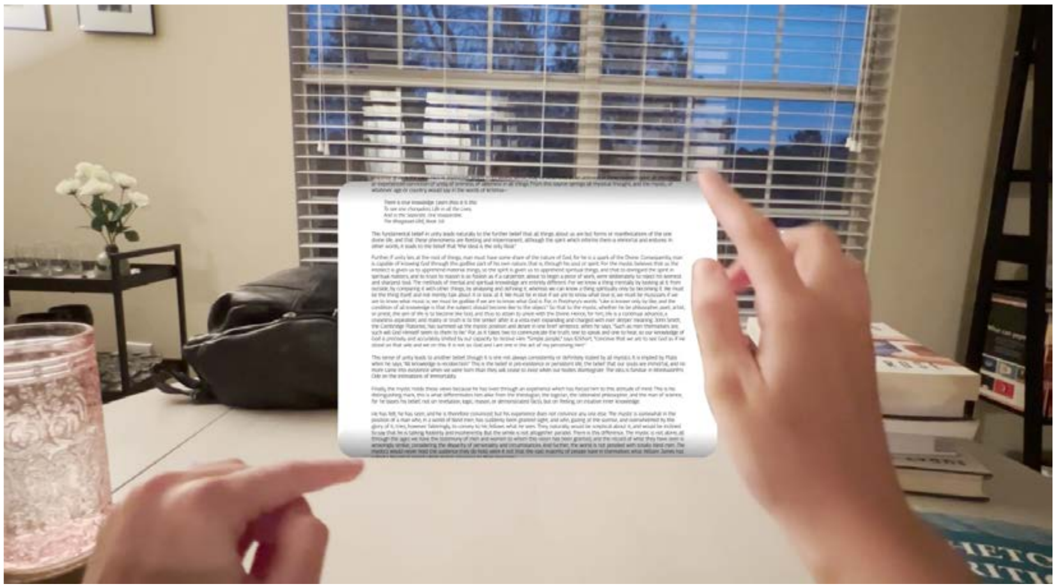


Figure 4.1.4 Rescaling the perimeter of the background.

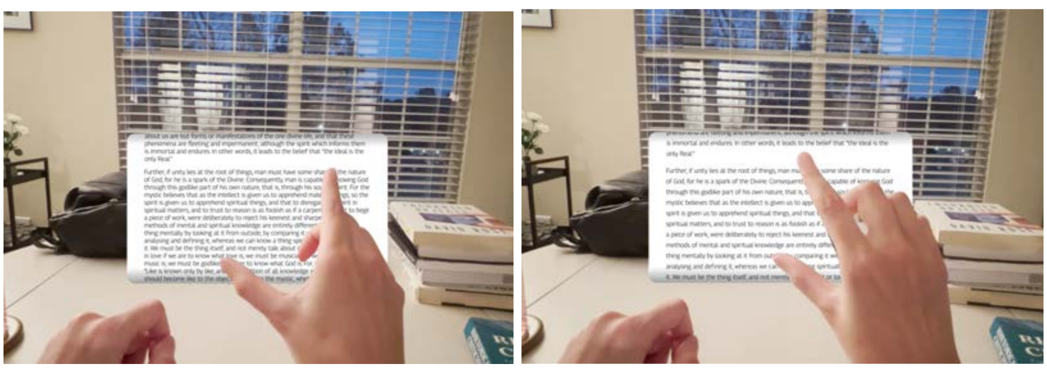


Figure 4.1.5 Thumb and forefinger spreading apart diagonally to increase font size (Left); Thumb and forefinger spreading apart vertically to increase line height.

Dividing the text into paragraphs is an additional exploration not shown above (Figure 4.1.6). Paper is an arbitrary size in a digital view, but a paragraph is a more purposeful division of the text initiated by the user. With the text divided into paragraphs, the spatial arrangement relies on depth to help the user understand where she is within the written text. Swiping up with the hand or a matching vocal cue that brings the next paragraph up might be too repetitive to be effective.

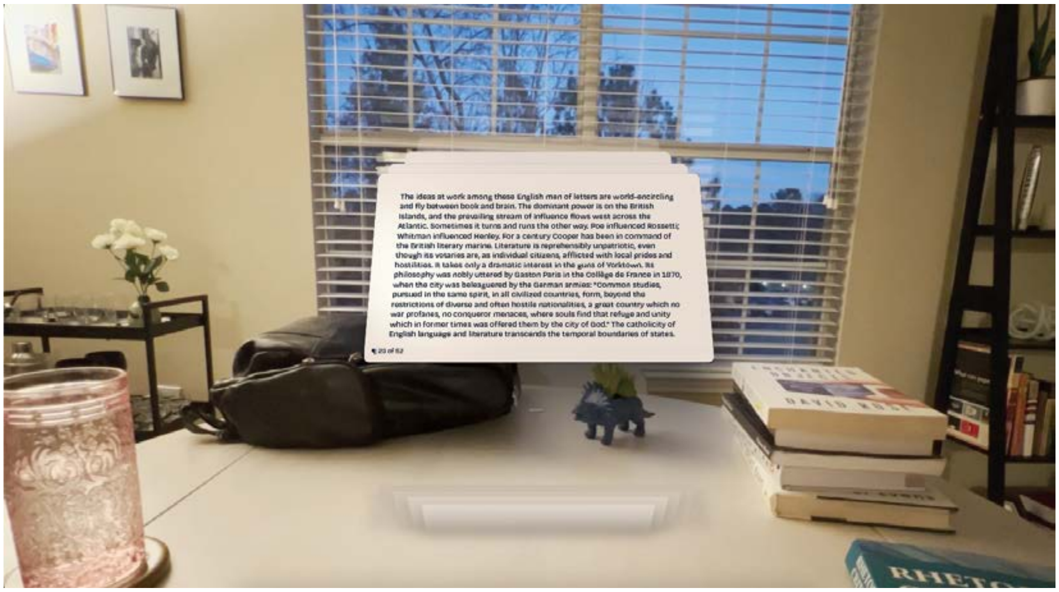


Figure 4.1.6 Dividing text by paragraphs.

ORAL COMMUNICATION WITH THE MACHINE

These ARSG interaction designs show collaboration with the machine. Today, many companies try to humanize devices and name them. In this scenario, the user recognizes the machine is a smart, non-human collaborator and communicates with the glasses in both human and nonhuman ways. As a collaborator, the user speaks terms such as "we" and "let's" when communicating about tasks. For simpler actions, ones that are activated with keyboard commands or shortcuts today, the user engages with non-lingual clicking noises. For example, the user verbalizes a "tsk" sound to create a new paragraph.

To grab the machine's attention, the user double clicks her tongue. Once the machine is listening, she can state a command. Likewise, the user double clicks her tongue and couples the sound with a hand gesture to open the home screen. In contrast, the closing function is a gestural interaction. For example, to close a paper, the user closes a fist in front of the object and lifts it upwards (Figure 4.1.7). This gesture automatically saves the written text to the Cloud[1].

Spoken text can take on the nature of the author's voice through a deepfake. A deepfake is synthetically generated content—such as audio, video, text, or image—created by machine learning to mimic the likeness of a person or theme. In this study, the deepfake is audio created by the vast amount of data of each user's voice. Deep fakes could be shared across

[1] The Cloud is a functionality metaphor that is flexible enough to still have value and meaning in this future scenario.

a user's social network for others to hear their messages or comments in that user's voice. If the author's voice is unknown, the voice output might be generic. The user double taps the area of the paper she wants to be read, and the machine responds by reading it in the author's voice (the user in this case) (Figure 4.1.8). The text that is spoken slightly hovers over the background and an indicator informs the user the location of the text that is being read.

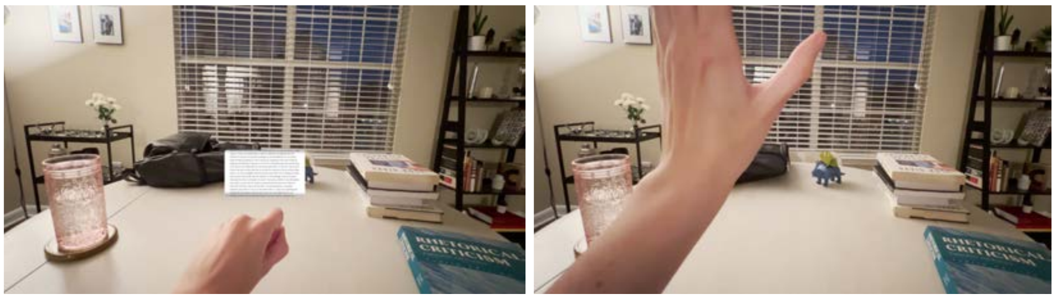


Figure 4.1.7 The paper scales down with a clenched fist as feedback to the user (Left); The paper ascends to the Cloud when the fist opens up towards the sky (Right).

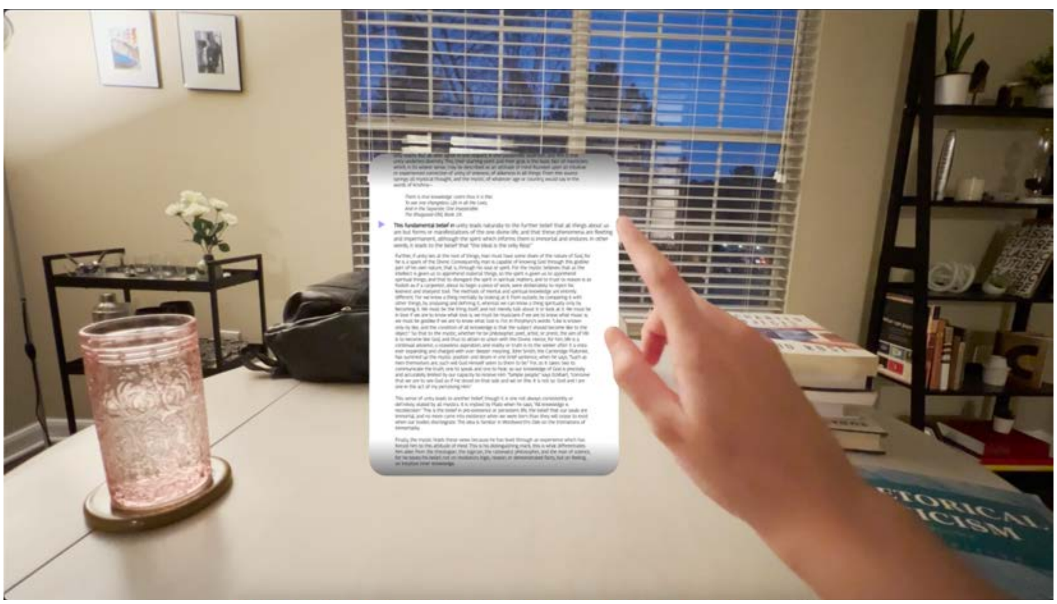


Figure 4.1.8 User double taps a paragraph for the machine to begin reading it aloud. The purple indicator shows which line the machine is reading and the text appears in bold as the machine reads the words.

DICTATING AND REFERENCING A TEXT

To begin dictating, the user places her forefinger where the text should start and vocalizes a double "tuh" sound. As she dictates, the machine records the spoken words as text and suggests grammar edits that can be accepted or declined (Figure 4.1.9). As the user edits text, the original transcript, recorded in a separate format, can be referenced by pulling from the paper with both forefingers (Figure 4.1.10). The transcript acts as a version history showing where edits were made and how the text has been rearranged. The transcription has inflection and tone encoded into its presentation to give the user an idea of how words were said without having to relisten to the recording. The inflection and tone are visualized through the use of variable type. Words or portions of words are italicized, bolded, extended, or compressed based on criteria such as the volume, inflection, tone, time elapsed, and other characteristics of her voice when the word was vocalized.

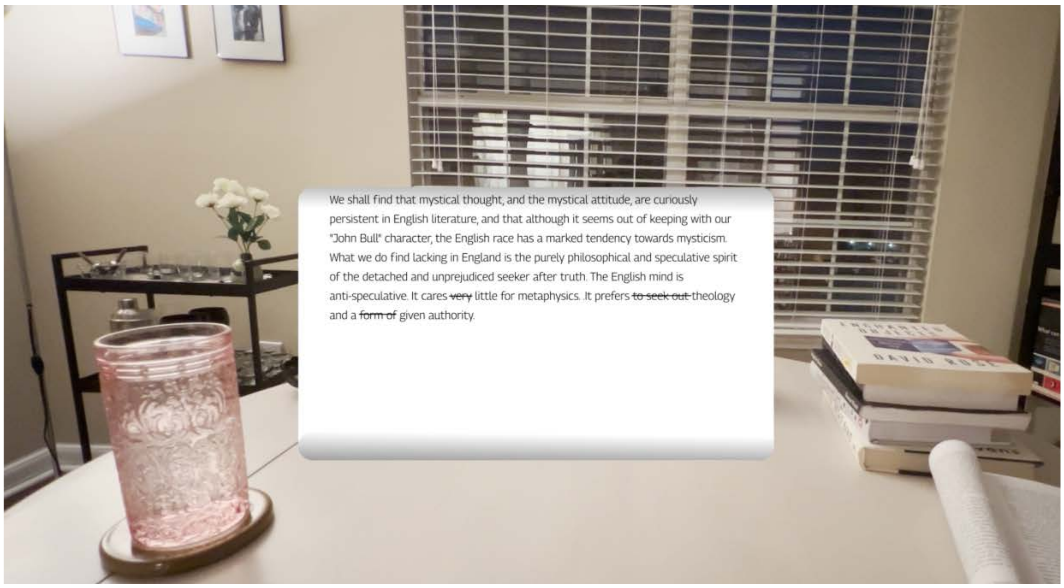


Figure 4.1.9 As the user dictates, the machine auto-suggests grammar edits.

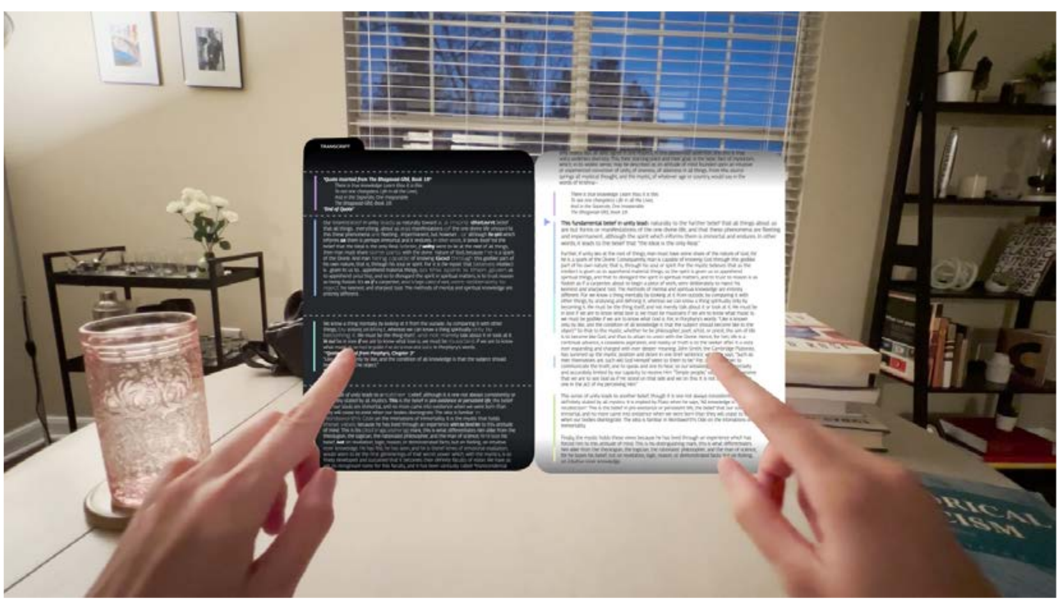


Figure 4.1.10 Pulling the transcript apart from, or out of, the edited text.

To listen to a colleague's comment, she taps on the comment and the deep fake of the colleague’s voice speaks (Figure 4.1.11). To add a comment, she pinches and pulls in the location of the text that she wants to comment on to create a connection between the text and her comment. A double “tuh” clicking sound indicates that she wants to record her comment. Similarly, connections can be made between reference materials and the essay she is writing. For example scenario below, the user opens the work she wants to reference and, once the user finds a spot to quote, she drags her finger over the text to highlight it, then pulls the quote over to the writing area (Figure 4.1.12). The quote is copied over, cited, and a connection link is created between the writing and the source.

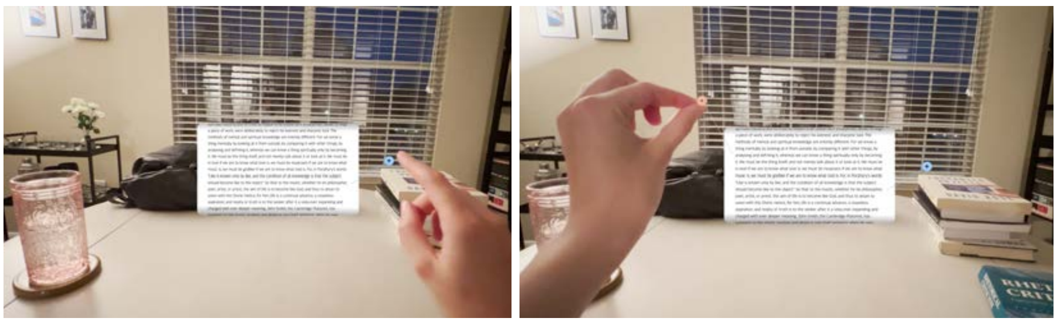


Figure 4.1.11 User taps on the comment button to hear a deep fake of the professor’s comments (Left); The user pinches and pulls a part of the text to create a voice comment (Right).

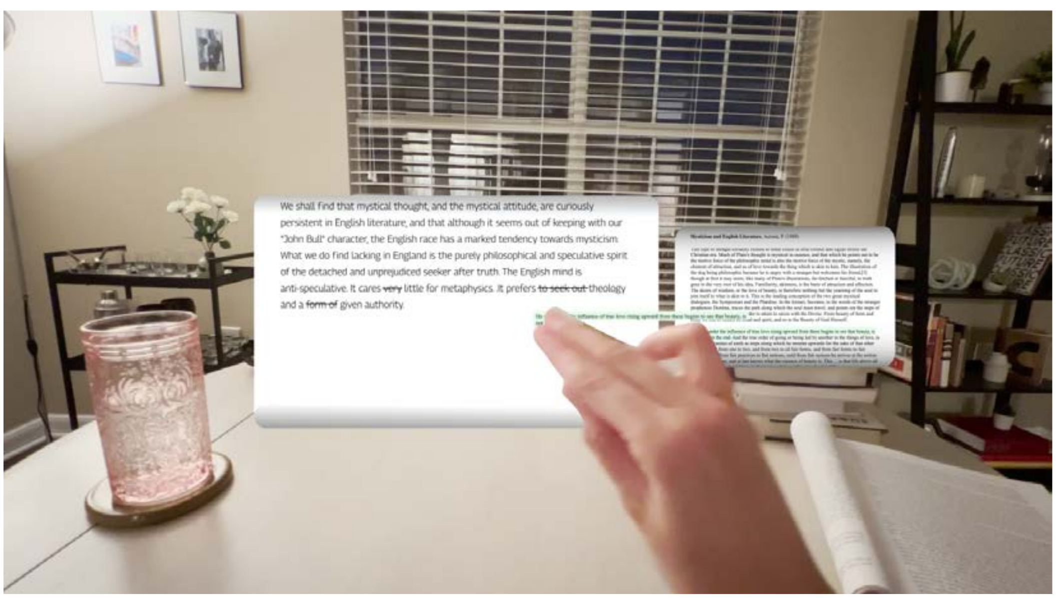


Figure 4.1.12 User pulls over a quote from a reference into her paper.

NON-SMART TO SMART OBJECT INTERACTION

Not everything will be digital or digitized in the future, but because ARSG offers a large field of view, non-smart objects can seemingly become smart through various technologies. For example, image-text recognition might allow a user to highlight and copy text from a physical textbook over to a digital interface (Figure 4.1.13), or have the textbook read aloud. Similarly, a book with a QR code or digital supplemental information could be accessed through a flicking gesture (Figure 4.1.14). The QR code could transform into a 3D object that can be flicked to be activated and opened within the ARSG interface.

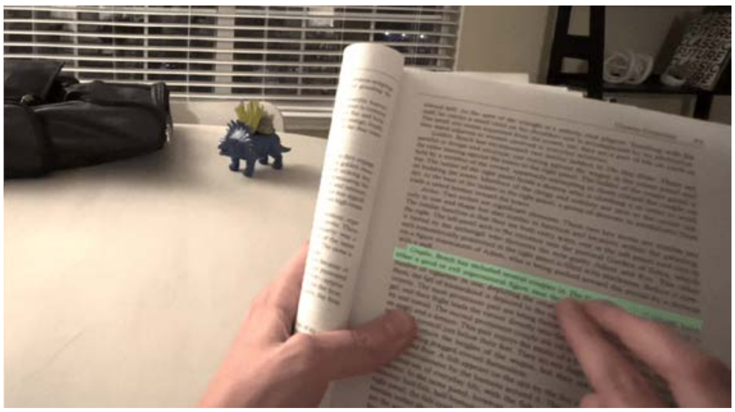


Figure 4.1.13 Image-text recognition with a physical textbook allows user to copy quotes

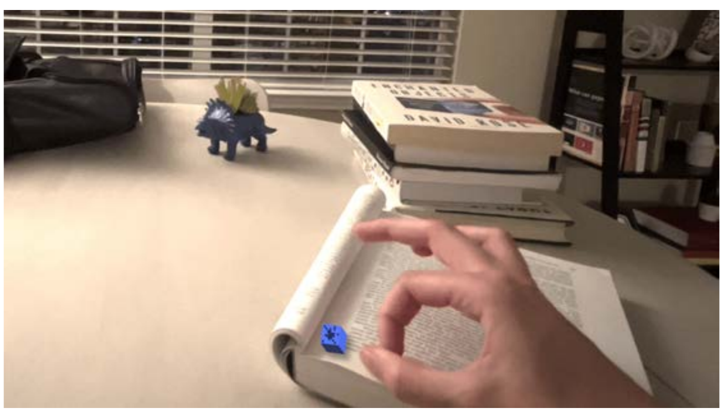


Figure 4.1.14 A 3D QR code with supplementary information is flicked by the user to “send” the digital information to the glasses

Desks can be messy with a lot of physical material. Digital files might be able to form a mess as well, or at least be visually in proximity to other papers lying about on the desk. Figure 4.1.15 shows a digital file being placed onto the table by the user closing her fist and positioning her hand downward.

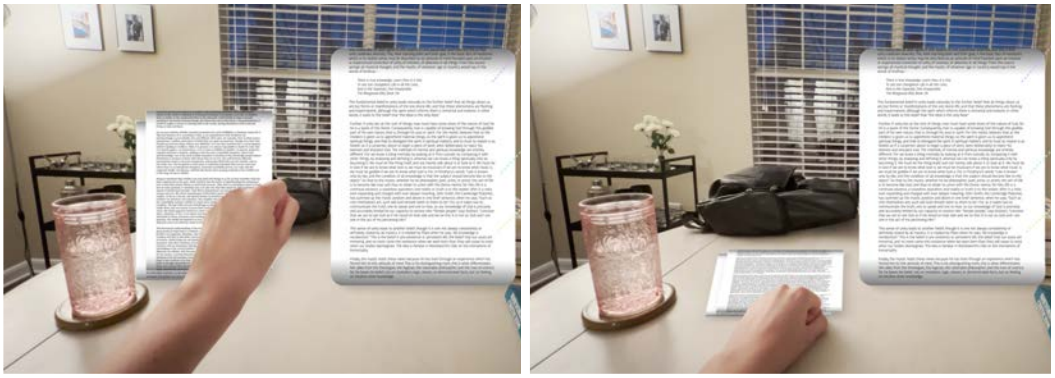


Figure 4.1.15 User clenches the papers in her hand and places them onto the table

CONTEXTUAL	ACCESSIVE	COLLABORATIVE
Dictating	Non-smart to smart object communicating	Annotating
Displaying text	Researching	Commenting
Recognizing vocal inflection	Sourcing + citing	Co-authoring

Table 4.1.1 Exploration examples relating to the Investigation Model

INTERACTION	MEANING / MACHINE ACTION
Tongue click + unspecified hand gesture	Opens visual homescreen (not prototyped)
Tongue click only	Glasses listens to you / your command
Finger held down in place + vocal double “tuh”	Starts recording and transcribes the dictation
Double tap	Machine reads text (double tap again to stop)
Vocal “tsk”	Enters down to a new paragraph
Pinch, drag and hold + double vocal “tuh”	Dictate to add a voice comment
Left / right full hand swipe	Moves through items organized left / right
Open hands come together	Collapses papers into infinite scroll
2 forefingers spread apart	Separates transcript from the edited text

Flick	Sends information from non-smart device to ARSG
Thumb and pointer spread diagonally apart	Scales the text
Thumb and pointer spread vertically apart	Increases line height of text
Fist grab and placed up	Closes and save items to the Cloud
Fist grab and placed down	Places the object onto the surface
Flat vertical hand flip	Reorders content

Table 4.1.2 Gestural and oral interaction explorations from Study 1

4.2. Creative, Visual-Based Tasks

My second investigation explores how to create things visually in an open and more active environment using Augmented Reality Smart Glasses. In contrast to writing academic prose, the subject of Study 1, the aims of the tasks are of an expressive nature, and the results are dimensional, specifically thinking through virtual making.

STUDY 2 PERSONA AND SCENARIO:

Rose is a 45-year-old poet and artist. She has used AR in her work since the start of her art practice back in the 2010s. Her current work is created through AR smart glasses and a stylus at times. She places her pieces around spaces for people to find and interact with—creating artistic experiences of text and visuals that complement a given location's spatiality.

Rose does not believe in digital ownership and that digital creation should be part of a shared and open community, acting as an extension of the physical world. At the start of a project, Rose finds a place to understand more deeply and create within. However, most of the actual creation time, is spent iterating on ideas and thinking about the particular mindsets Rose experiences. She has made an algorithmic typography tool that responds specifically to her voice as she writes poetry, which then is manipulated using voice inflection.

PROTOTYPED VIDEO

The video is linked below or can be viewed on YouTube at: https://youtu.be/Jp39JUCPfFg

THE BODY IN CONTEXT

The artist plays an active role. Her creative actions are often embodied, moving her entire body to create sizable strokes through the environment. The supplemental interfaces that I explored throughout this study act as peripheral attachments to the artist's body that are brought into view at any time through gestures. One such interface is a variable toolbar that sits at her non-dominant (left, in this case), shoulder. To draw the interface into view, she moves her flat left hand from her left shoulder to the center of view with the palm facing her (Figure 4.2.1). She can then manipulate the tools to her liking and push the interface back out of view to begin creating. The tools act as extensions of her freely moving body without disrupting her view.

Artists often evolve a visual style that reprises elements or repeats techniques. The functionality metaphor of a *bin* gathers such elements into one place. The virtual bin sits at her feet (much in the way that a "cloud" hovers above), and can be brought up by raising an open palm from her belt to the center of the frame. From this bin she can retrieve saved or frequently used elements and styles.

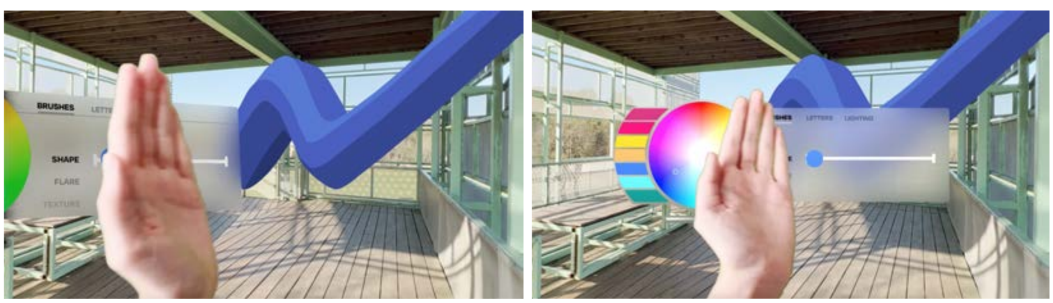


Figure 4.2.1 User pulls the toolbar out of the periphery and into the center of view

VARIABILE FUNCTIONALITY

Design applications such as Adobe Illustrator or Photoshop are riddled with icons representing different types of tools. Icons, as abstract pictographs for interface actions (with some based on outdated technology such as the pen tool), require users to memorize corresponding functions. Icons as a form of representation ultimately conceal the information that they are meant to convey (Raskin, 2000). The Adobe applications are quite complex and icons can lessen space usage in an interface, depending upon how many are in use at a given moment. This study explores a sliding functionality metaphor as a replacement for brush, pen tool, shape and text icons.

The same sliding toolbar functions to manipulate the roundness, shape and textures of virtual brushes in a similar manner to how variable type works currently. I studied only three facets of a brush stroke, but the concept is transferable to any number of variables. The *shape* slider changes the roundness of the shape from a circle to a rectangle (Figure 4.2.2). The *flare* slider puckers the shape when pulled to the right, and bloats it when pulled to the left (Figure 4.2.3). When the user rotates her hand to the left, anchor points are added to the shape to create more instances of puckering, and the opposite is done when she rotates her hand towards the left. The *texture* slider changes the material texture of the shape (Figure 4.2.4). The view visually reveals the change in real time, utilizing textual identifiers for each facet rather than menus and lists of iconography.

The color wheel is ever present with a side palette of recently used swatches. The color of the brush shape automatically updates when the user selects and uses a color from the wheel. This transparency in the interface supports real time visualizations of the form in the midst of creation.

Figure 4.2.2 User is transforming the roundness by sliding the shape tool and creates a rounded rectangle

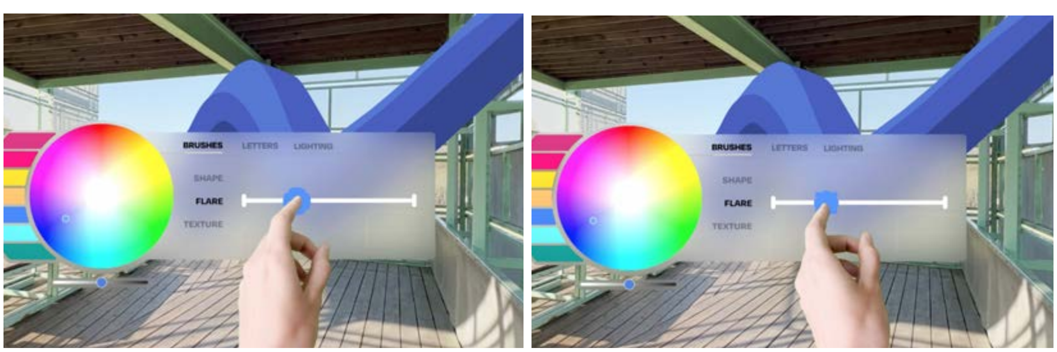


Figure 4.2.3 User slides rounded rectangle to the left which adds a bloating effect (Left) and the user rotates her hand to the left to create more areas of bloat (Right)

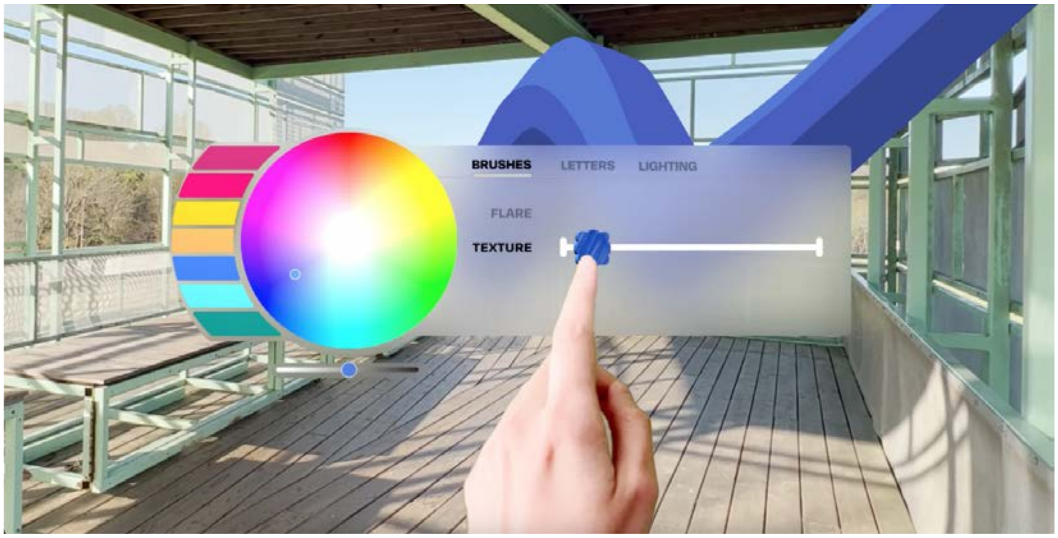


Figure 4.2.4 The user changes the texture of her shape by sliding it across the slider

CREATING SHAPE AND DIMENSIONALITY

The user holds a stylus in her right hand to better control the form of her shapes. To increase the stroke weight, she repeatedly moves her forefinger outward over the stylus; to reduce it she moves her forefinger inward. The more repeated motions, the larger or smaller the stroke gets. Once she draws the outline of a shape, specific gestures produce different types of shapes. For example, if she holds a fist, the shape fills with a solid color. Next, she can pull the shape in any direction to extrude it (Figure 4.2.5). A cupped hand over the shape while moving keeps the shape outlined while pulling in any direction that extrudes the shape (Figure 4.2.6). To revolve a shape, she cups her hand and twists it around (Figure 4.2.7). [Note: Table 4.2.2 is updated to reflect a change from the prototype. A revolved shape can be created with a fist *or* cupped hand that rotates 180 degrees.] To flatten any shape, she presses a flat hand on top of it (Figure 4.2.8).

The fist as a metaphor for grabbing has been used in Study 1 and 2. A cupped hand and flat hand could become useful gestures within a spectrum of actions starting from a fist. The fist suggests a closed or absolute action, whereas a cupped hand is more open yet can signify containment. A flat hand can represent halting or releasing the action, as investigated in Study 1. Here it represents flattening.

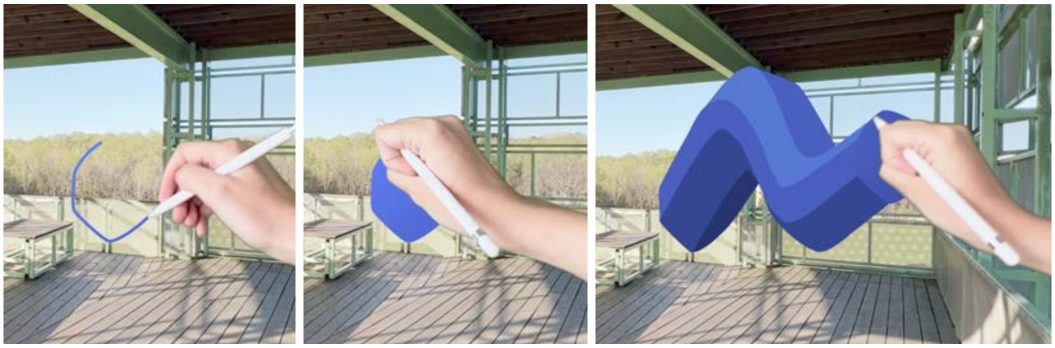


Figure 4.2.5 Drawing a shape with a stylus (Left), filling the shape solid with a closed fist (Right), and extruding the filled shape with a closed fist by pulling it in the preferred path of the shape.

Figure 4.2.6 Extruding an outlined object with a cupped hand.

Figure 4.2.7 Cupped hand rotating to create a revolved shape [Note: the gestures have been updated allowing a fist to create a solid revolution and a cupped hand to create a hollow revolution.]

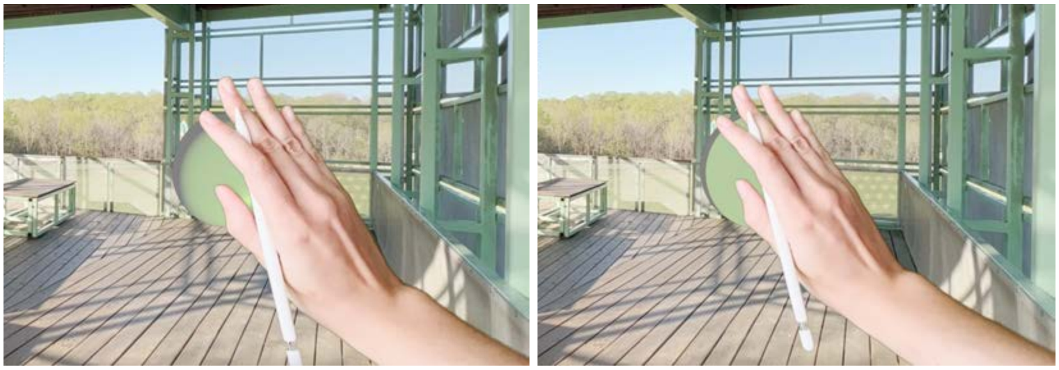


Figure 4.2.8 Flat hand continually pressing the shape deflates it (Left) and eventually leaves it flattened (Right)

PRODUCING AND MANIPULATING TYPE

Variable type is used to express the tone and inflection of the artist's voice when speaking verses. She has written a specific algorithm for how each type variable—such as the width, height, and weight–responds to the volume, tone, and inflection of her words, and the speed of utterance (Figure 4.2.9). Once the machine visualizes her words, she can transform the letters or words. The type becomes a specific personal reflection of the artist at that specific moment in time. This feature reinforces the iterative process of creating endless manifestations of voice-to-text output. This way of capturing the character of the voice within variable typography can be applied to other forms of communication where voice input can be encoded with meaning through visualization. The concept also presents possibilities for visualized inflection that is generalized or personal to each user.

Figure 4.2.9 Type responding to the user's tone and inflection of voice

Handwritten input or lettering is possible when designed fonts are not desired for creative expression. In certain situations, the artist might not need to create an entire alphabet by hand. In Figure 4.2.10, the artist creates four defining letterforms, and the machine then creates the remaining glyphs needed. This study explores options for text-based outputs through 1) variable type encoded with inflection, 2) lettering combined with machine learning to produce a font, and 3) free form drawing to produce type.

Figure 4.2.10 The user draws a few foundational letters of the font to be created (Left) and the machine creates a font that can be used based on the letterform designs (Right)

The extrusion of the letters is handled in the same way that shapes are extruded: with the artist's fist pulling the letters in a direction (Figure 4.2.11). To move a piece of type, she uses a flat hand to underline the portion she wants to move and then relocates it to the preferred position (Figure 4.2.12). The flat hand gesture acts as a way of selecting a portion of an item, in contrast to a fist that grabs the entire object.

Figure 4.2.11 A fist extrudes the letters in the same way it extrudes a shape

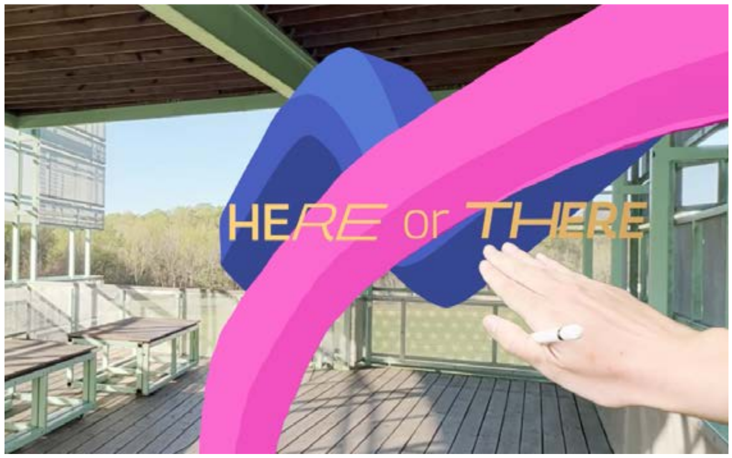


Figure 4.2.12 The user underlines the letters she wants to move

CONTEXTUAL	COMMUNICATIVE	COLLABORATIVE
Drawing	Public digital art	Shared digital ownership
Manipulating shapes	Algorithmic type	Machine learning type creation
Visualizing verbal expressions	Expression of thought	Output preview through variability

Table 4.2.1 Exploration examples relating to the Investigation Model

INTERACTION	MEANING / MACHINE ACTION
Closed fist over drawn outline	Fills the shape solid
Fist pulling a shape along a path	Extrudes a solid shape along the path
Cupped hand over drawn shape	Keeps the shape outlined while being manipulated
Cupped hand pulling a shape along a path	Extrudes a hollow shape along the path
Non-dominant flat hand moving from nondominant shoulder to the center of frame	Slides the toolbar interface from the periphery to the center of frame
Forefinger as direct manipulation	Fingertip acts as a cursor
Double tongue click; followed by another double tongue click	Records and visualizes the spoken words; ends the recording
Flat hand underlining or encircling	Selecting a portion of an object

Pinched hand over an object	Selects whole object
Two cupped hands over an object	Molds the object's shape
Fist turning 180° over an outlined shape	Creates a solid revolved shape
Cupped hand turning 180° over an outlined shape	Creates a hollow revolved shape
Flat hand pressing dimensional object	Flattens the object

Table 4.2.2 Gestural and oral interaction explorations from Study 2 with minor edits from prototyped visual

4.3. Synchronous and Asynchronous Communication Tasks

Study 3 explores communication within an ARSG environment from the perspective of a curator and artist. This study investigates communication channel restructuring and the exchange of expression between input and output sources.

STUDY 3 PERSONA AND SCENARIO:

Tan is a curator for a state art museum. He works closely with artists as they install their exhibitions. Today, Tan is meeting with an artist, Rose, to discuss the location of her upcoming AR sculpture exhibit in the Art Park. He regularly calls and messages artists, co-workers, and donors as he plans and organizes exhibitions. Tan uses his AR smart glasses for both work and personal activities. He prefers talking to people synchronously, but he usually dictates the message if he must message someone. Rose also has ARSG and uses them to create and communicate with exhibition collaborators, like Tan. She primarily dictates asynchronous messages to others because of her busy schedule but prefers to send the messages visually, usually through typography or video.

PROTOTYPED VIDEO

The video is linked below or can be viewed on YouTube at: https://youtu.be/pXjZHoEr-Gg

RESTRUCTURING COMMUNICATION METHODS

Currently, electronic and cellular communication forms—calls, texts, or video chats—are distinct modes within our devices and are usually separated into discrete applications, such as the messages app for text or phone app for calls. While there are other forms of communication, this study focuses only on these three. I categorized each form of communication into its fundamental function: text as type, call as voice, and FaceTime (or other video calling apps) as video (Figure 4.3.1). Newer cross-media features in each category are beginning to merge the categories. For example, voice to text transcribes verbal input into text output, voice messages are recorded messages sent through text, and so on. I'm projecting the communication channels will continue to merge and ultimately require a restructuring of categorization.

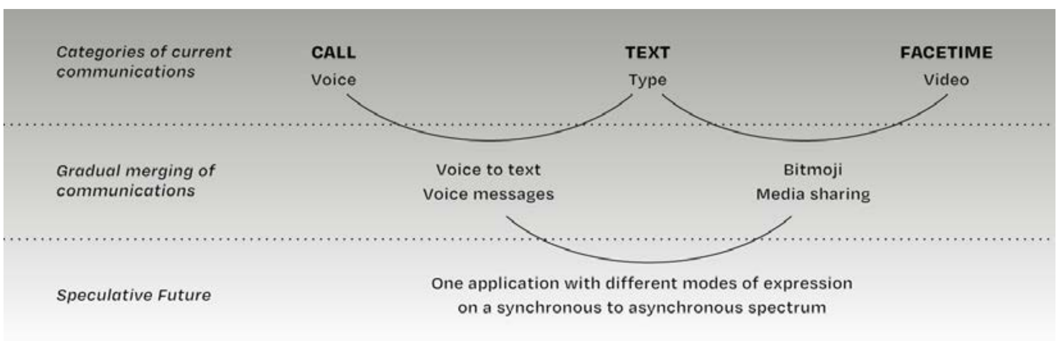


Figure 4.3.1 Gradual change in forms of communication between now and the speculative future scenario

In the plausible speculative future, the fundamental differences between the types of communication are not their input or output form, but rather whether or not the communication is synchronous or asynchronous. This study applies the functionality metaphors parallel communication (synchronous), or serial communication (asynchronous). In Figure 4.3.2, parallel communication is represented by horizontal, abstract parallel lines next to the communication style–voice, video, and text–and serial communication is represented by vertical lines, one on top of the other (Figure 4.3.2). The gestures necessary to engage in synchronous or asynchronous communication mimic the abstract representations. To begin a parallel communication, the user slides his hand to the right; and to start a serial message, the user slides his hand downwards (Figure 4.3.3). The lines from the triangle then follow the gestural path and open the corresponding options.

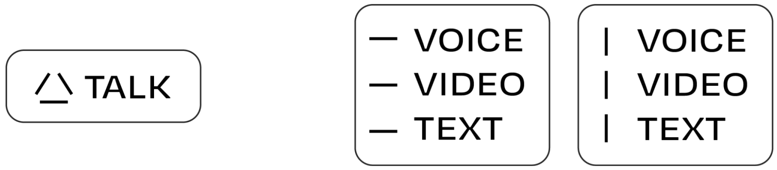


Figure 4.3.2 Generic 'Talk' button as an entryway to communicating (Left) with the triangle morphing into parallel communication (middle) or serial communication (right) depending on the user gesture

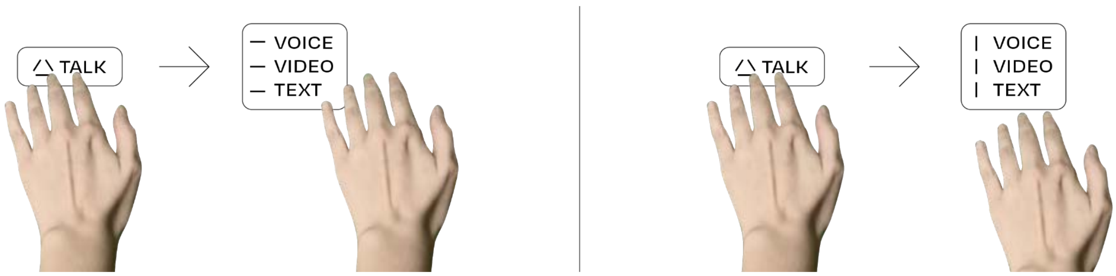


Figure 4.3.3 Horizontal gesture to create parallel communication (Left) and vertical gesture to create serial communication (Right) as mimicked by the line icons

INVERSIONS OF MEDIA

Listening to another's voice involves hearing the tone or inflection. Text and other non-vocal communication are expected to carry the same amount of expression in the plausible future. As discussed above, the input communication forms are voice, text, and video, but the output does not have to match the input. The recipient can receive an output different from the input, making outputs malleable (Figure 4.3.4). For example, a user may dictate a message (input) that is then sent as a textual video (output) to the receiver, or a user may type a message (input) that is then sent as a voice message using deepfake technology (output). GANs (Generative Adversarial Networks) can create deepfakes, or facsimiles of the sender's voice, through a believable adversarial learning model on the user's voice (Goodfellow, 2014). Choice adds flexibility for the user and makes more forms of expression readily available to better suit the message's tone.

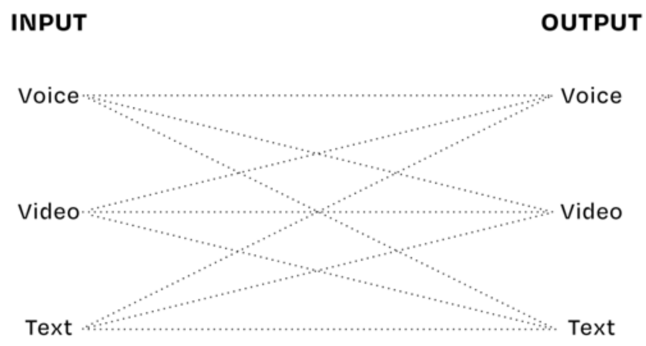


Figure 4.3.4 Forms of input and its flexible choice of output

Figure 4.3.5 shows a user expressing congratulations through the multimedia input of voice and video combined. Rose receives a serial message from her friend Sarah telling her, “Congrats on the exhibit and good luck today,” combined with a video of confetti falling. Rose verbally tells her ARSG to send a text message saying, “Thanks, Sarah,” and make the statement's textual visualization appear “excited” to enrich and enliven her thank you message. The textual message appears in all-capital, large letters with two exclamation points. The implications of inverting or making media more malleable extend beyond communication styles. Document file structures could, for example, have more extreme exporting and outputting abilities. In the future, systems are not dependent on embedding media into existing text documents. Moldable file outputs coexist together rather than as one embedded into another.

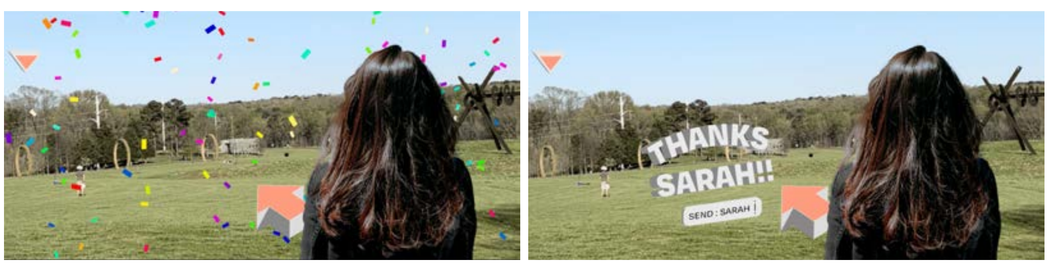


Figure 4.3.5 A serial voice message and confetti are sent to Rose (Left) and Rose responds with a serial textual message based on her voice input (Right)

READING PATHS AS A METAPHOR

Left to right reading progression is common in most Western cultures. Left, then, might be the starting point and right the endpoint. Left to right gestural movements might represent: moving forward, doing an action, or the passage of time. A right to left movement might mean going backward, ending an action, erasing, or undoing. As a user progresses through an interface flow, left to right gesture indicates a continuation of an action, as seen in Figure 4.3.6, while a right to left gesture indicates the need to delete, restate or go back. These gestures might replace a back button and function as a universal undo action in the ARSG environment.

Figure 4.3.6 Moving from left to right (mirrored image) through a progressive flow

ORAL COMMUNICATION AND JUST IN TIME INTERACTIONS

Oral input is a rich form of data for machine learning software. If the system is permitted to listen, a user who states or implies potential following actions in conversations with others would trigger suggestions based on the keywords. There are many privacy concerns related to this idea, but the data can provide rich, *'just in time'* recommendations, which relieves the user of directly commanding an interface. For example, when Tan states to Rose that he will send the meeting location, the interface responds and proposes the same action in a question, which gives Tan the power to agree or disagree with the action (Figure 4.3.7). Tan navigates around the 3D map to find the meeting location, pinches and pulls upward at the location, and makes a popping sound to complete the action. The ARSG recognizes a location has been selected and gives the option of sending the proposed meeting location to Rose. A similar interaction might apply when sharing information or media with others in addition to locations, such as photos or social media content.

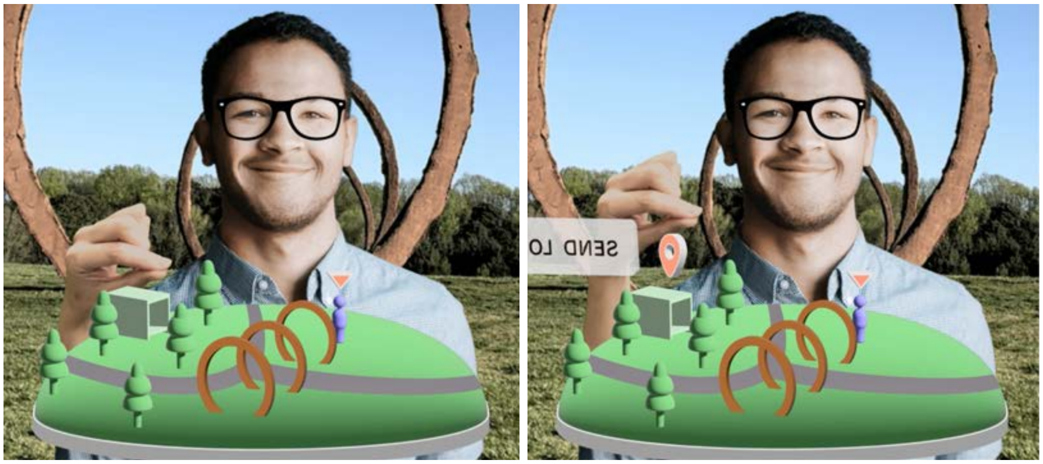


Figure 4.3.7 Pinching and pulling upward to place a location marker

CONTEXTUAL	ACCESSIVE	COMMUNICATIVE
Sending location	Synchronous communication	Expressive output forms
AR directions	Asynchronous communication	Calling / Messaging

Table 4.3.1 Exploration examples relating to the Investigation Model

INTERACTION	MEANING / MACHINE ACTION
Sliding flat hand left to right over talk interface	Opens parallel/synchronous communication options
Sliding flat hand top to button over talk button	Opens serial/asynchronous communication options
Pinch and pull over object on map	Creates a map marker
Receiving a location	Arrows navigate user to location / Caret in the sky above end location
Asynchronous voice, type, or video input	Output to asynchronous voice, type or video
Synchronous voice, type, or video input	Receiver outputs to synchronous voice, type, or video input

Table 4.3.2 Gestural and oral interaction explorations from Study 3

Discussion

5.1. Design Principles

IMPLEMENT FLEXIBLE FUNCTIONALITY METAPHORS

Metaphors are central to communication, but the type of metaphor used can be curated. Since functionality metaphors are the overarching concept of what the *thing* does, each functionality metaphor should be flexible enough to allow for a scalable system of interface and interaction metaphors to complement the functionality without breaking the metaphor. Avoid visualizing concrete interface metaphors, and instead, consider how to create mental models or patterns for users to understand.

CREATE CONSISTENT SPECTRUMS OF GESTURAL METAPHORS

Differentiated gestural actions are limited. Many gestures are continuations from one gesture to another—the gestures between an open hand and a fist, or from one finger up to all five. Across the ARSG platform, a gesture continuum should be designed consistently for related actions.

CHALLENGE INPUT FORMS

Should language and dictation become the primary input mode for future ARSG interfaces, consider creating oral shortcuts for simple tasks—tasks congruent with what keyboard shortcuts offer users today. The oral shortcuts that I explored in this investigation were mainly different clicking noises of the mouth, but other possibilities need to be explored.

EMBRACE THE PHYSICAL WORLD

Embrace the mess of the physical world and include non-smart objects into the interface to enrich the user's physical presence in the world. ARSG can be a solution for turning our attention back from the device at hand to our physical surroundings again. AR Smart Glasses offer a full range of view to bring the digital and the physical worlds together, as well as allow information to straddle the physical world and the interface.

UTILIZE AR AFFORDANCES AND EXPLORE NON-INTUITIVE EXPERIENCES

ARSG interface design and flat GUIs are different technological paradigms. The former is primed for novel explorations. Many of the ARSG interfaces currently mimic flat, screen-based interfaces in an AR environment. But an ARSG interface can do much more. Instead of point-and-tap interactions, consider embodied, gestural experiences. Instead of flat menus for navigation and information concealed behind iconography, consider ways of spatially displaying information that might not be bound to containers. Intuitive design is a familiar design, and most consumers are not familiar with ARSG interface design . Utilize AR's affordances and create non-intuitive designs because more suitable approaches have yet to be discovered.

5.2. Future Work

FEASIBILITY OF GESTURAL AND ORAL INTERACTIONS

These studies are only the tip of the iceberg (metaphor intended) for how gestural and oral interactions could function within an ARSG environment. Deeper exploration of more complex and thorough application usage of a gestural/oral interaction system is required, as well as documentation of all of the options available. Ultimately, it should be concluded whether a gestural and oral interaction system is feasible for ARSG interfaces.

TYPOGRAPHIC PRINCIPLES FOR ARSG

There are many typographic unknowns in the world of AR. Legibility and readability standards are perhaps the most important principles to study immediately. Beyond a surface level look into readability or color contrast between the type and background, a larger question of how AR changes the way users read emerges. When should type follow the user's head movements versus stay in place? How does oral input change the way type is displayed in an ARSG interface? Is tone and inflection worth encoding into voice-to-type messages?

SHARED DIGITAL OWNERSHIP

As an ARSG interface overlays its information onto the full range of a user’s view, it comes with a sense of universality and shared experience. Which parts of an AR system are shared with every user and which parts are private? For example, should every user have to see a piece of AR public art? Should a user be able to leave a digital trace of himself for another to find? A larger question of what parts of an omnipresent system are shared, which can be turned on or off by the user, and which are private to each user needs to be explored.

ALTERNATIVES TO FLAT GUI INTERACTIVE ELEMENTS

As stated in the Design Principles section, ARSG interface design is a different paradigm with distinct affordances compared to GUIs. Desktop and mobile interfaces rely on flat buttons, links, and iconography to visually navigate, usually designed in space-saving ways due to the screen boundaries. Alternatives to linking and displaying information spatially need to be thoroughly and fully investigated.

ARSG HOME SCREEN

This investigation began as a first step in understanding how an ARSG interaction system could work across different user tasks. The next step is to take the learnings from this study—and many other ARSG studies—and create an ARSG "home screen" interface and interaction design system.

5.3. Conclusion

This investigation looks at possible interaction and functionality metaphors for Augmented Reality Smart Glasses technology within a near-future scenario. The studies explore alternatives to concrete HCI metaphors–the current flat interface paradigm–and familiar WIMP (Windows, Icons, Menus, Pointer) interactions to create more spatial and flexible metaphoric interactions for ARSG interfaces. Principles surrounding flexible functionality metaphors, spectrums of lingual and non-lingual gestural interactions and oral input, and variable, transparent interfaces for ARSG, were derived from the investigation. The research suggests gestural and oral interactions mapped from sensorimotor and cultural knowledge might replace the need for user interactions mapped from familiar WIMP interactions. Flexible functionality metaphors, such as the parallel/serial communication metaphor, might encourage users to create patterns or mental models of the abstracted representation rather than map concrete functionality onto a digital object.

ARSG interface design is a new design paradigm with implications that require further research from multidisciplinary experts in the field. Researchers might continue to explore plausible future scenarios to advance practices for augmented spatial lens-based design. Although the findings in this investigation are not final solutions, the possibilities can be used to further develop modes of ARSG interaction.

5.4 Program Statement on the Master of Graphic Design Final Project

This document details a final project, which in design is commonly referred to as a graduate "thesis" at North Carolina State University. The work was defined in a 3-credit course in a fall semester, and executed in a 6-credit course in the following spring semester. The Master of Graphic Design is a terminal professional degree with a research orientation, but like the MFA and MDes, it is not a primary research degree. This investigation is discovery-based. Cash (2018) describes the process of building scientific knowledge as a cycle between theory building and theory testing. The theory building mode includes (1) discovery and description, (2) definition of variables and limitation of domain, and (3) relationship building (pp. 88–89). This investigation is restricted to the theory building mode. The theory testing mode includes (4) prediction, testing, and validation, and (5) extension and refinement (p. 89). While experts may have been consulted, this investigation does not entail any testing with human subjects, and it does not endeavor to prove anything; all assertions are tentative and speculative.

See: Cash, P. J. (2018). Developing theory-driven design research. *Design Studies, 56,* 84–119.

References

Agarawala, A. (2016) BumpTop – A Multi Touch 3D Physics Desktop [Photograph]. GitHub. https://github.com/bumptop/BumpTop/wiki

Agarawala, A., & Balakrishnan, R. (2006). Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 1283-1292).

Alden, D. G., Daniels, R. W., & Kanarick, A. F. (1972). Keyboard design and operation: A review of the major issues. Human Factors, 14(4), 275-293.

Banky, G. P. and Blicblau,A. S. (2017), Doing it with a glass onion: Investigating affordances provided by a head-mounted augmented reality immersive device for the real-time online supervision of experimental learning. 2017 IEEE 3rd International Conference on Engineering Technologies and Social Sciences (ICETSS), pp. 1-6, doi: 10.1109/ICETSS.2017.8324168.

Blackler, A., & Popovic, V. (2015). Towards intuitive interaction theory. Interact. Comput 27 (3). 203-209.

Blackwell, A. F. (2006). The reification of metaphor as a design tool. ACM Transactions on Computer-Human Interaction (TOCHI), 13(4), 490-530.

Candy, S., & Kornet, K. (2019). Turning foresight inside out: An introduction to ethnographic experiential futures. Journal of Futures Studies, 23(3), 3-22.

Dunne, A., & Raby, F. (2013). Speculative everything: design, fiction, and social dreaming. Cambridge, Massachusetts: The MIT Press.

Earnshaw R. (2018) Research and Development on Interfaces of the Future. In: Research and Development in Digital Media. SpringerBriefs in Computer Science. Springer, Cham.

Engelbart, D. (1962). Augmenting human intellect: A conceptual framework. Menlo Park, CA.

Fineman, B. (2004). Computers as people: human interaction metaphors in human computer interaction. The School of Design.

Gentner, D., & Nielsen, J. (1996). The anti-mac interface. Communications of the ACM, 39(8), 70-82.

Gibson, J. (1979). The Ecological Approach to Visual Perception. Houghton Mifflin Harcourt (HMH), Boston.

Glăveanu, V. (2012), What Can be Done with an Egg? Creativity, Material Objects, and the Theory of Affordances. J Creat Behav, 46: 192-208. https://doi-org.prox.lib.ncsu.edu/10.1002/jocb.13

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Hurtienne, J., & Blessing, L. (2007). Design for Intuitive Use-Testing image schema theory for user interface design. In DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007 (pp. 829-830).

Hurtienne, J., Klöckner, K., Diefenbach, S., Nass, C., & Maier, A. (2015). Designing with image schemas: resolving the tension between innovation, inclusion and intuitive use. Interacting with Computers, 27(3), 235-255.

Jetter, HC., Reiterer, H. & Geyer, F. Blended Interaction: understanding natural human–computer interaction in post-WIMP interactive spaces. Pers Ubiquit Comput 18, 1139–1158 (2014). https://doi.org/10.1007/s00779-013-0725-4

Johnson, M. (1987). The body in the mind: The bodily basis of meaning, imagination, and reason. Chicago: University of Chicago Press.

Khoury, G. R., & Simoff, S. J. (2004). Elastic metaphors: expanding the philosophy of interface design. In Computers and Philosophy Conference. Australian Computer Society Inc.

Kiseleva, J., Williams, K., Jiang, J., Hassan Awadallah, A., Crook, A. C., Zitouni, I., & Anastasakos, T. (2016, March). Understanding user satisfaction with intelligent assistants. In Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (pp. 121-130).

Kövecses, Z. (2020). An extended view of conceptual metaphor theory. Review of Cognitive Linguistics. Published under the auspices of the Spanish Cognitive Linguistics Association, 18(1), 112-130.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.

Lee, J. (2015). SpaceTop [Video]. Vimeo. https://vimeo.com/130447706?embedded=true&source=vimeo_logo&owner=2692098

Lee, J., Olwal, A., Ishii, H., & Boulanger, C. (2013, April). SpaceTop: integrating 2D and spatial 3D interactions in a see-through desktop environment. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 189-192).

Fineman, B. (2004). Computers as people: human interaction metaphors in human computer interaction. The School of Design.

MacKay, D. (2003). Dasher–an efficient keyboard alternative. Advances in Clinical Neuroscience and Rehabilitation, 3(2), 24.

Material Design. (n.d.). *Principles Section.* Retrieved November 17, 2021, from https://material.io/design/introduction

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W.H. Freeman.

Neisser, U. (1978) Perceiving, Anticipating, and Imagining. University of Minnesota Press, Minneapolis. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/185331.

Nielsen, J. (1993). Noncommand user interfaces. Communications of the ACM, 36(4), 83-99.

Norman, D. A. (1999). Affordance, conventions, and design. interactions, 6(3), 38-43.

Ong, W. J. (1982). Orality and literacy: The technologizing of the word. Taylor & Francis Group.

Peterson, M. (2021) Design Exploration as a Research Discovery Phase: Integrating the Graduate Design Thesis with Research in the Social Sciences. Dialectic, 3.1.

Raskin, J. (2000). The Humane Interface: New Directions for Designing Interactive Systems. Addison-Wesley Professional.

Rauschnabel, P. A. (2018). A conceptual uses & gratification framework on the use of augmented reality smart glasses. Augmented reality and virtual reality, 211-227.

Rauschnabel, P. A. (2018). Virtually enhancing the real world with holograms: An exploration of expected gratifications of using augmented reality smart glasses. Psychology & Marketing, 35(8), 557-572. https://doi.org/10.1002/mar.21106

Rorty, R., & Richard, R. (1989). *Contingency, irony, and solidarity*. Cambridge University Press, 152.

Sease, R. (2008). Metaphor's role in the information behavior of humans interacting with computers. Information technology and libraries, 27(4), 9-16.

Shin, D. (2021). Does augmented reality augment user affordance? the effect of technological characteristics on game behaviour. Behaviour & Information Technology, 1-17. https://doi.org/10.1080/0144929X.2021.1928286

Thacker, C., McCreight, E., Lampson, B., Sproull R., & Boggs, D. (1979) Alto: A personal computer. Xerox, Palo Alto Research Center.

Wilson, M. (2002). Six views of embodied cognition. Psychonomic bulletin & review, 9(4), 625-636.

Yuan, J. (2019) *Humane Section,* Mercury OS. Retrieved November 4, 2021 at https://www.mercuryos.com/