Discoverable Free Space Gesture Sets for Walk-Up-and-Use Interactions

by

Lavinia Andreea Danielescu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2018 by the
Graduate Supervisory Committee:

Erin A. Walker, Co-Chair
Winslow Burleson, Co-Chair
Kurt VanLehn
Anastasia Kuznetsov
Mary Lou Maher

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Advances in technology are fueling a movement toward ubiquity for beyond-the-desktop systems. Novel interaction modalities, such as free space or full body gestures are becoming more common, as demonstrated by the rise of systems such as the Microsoft Kinect. However, much of the interaction design research for such systems is still focused on desktop and touch interactions. Current thinking in free-space gestures are limited in capability and imagination and most gesture studies have not attempted to identify gestures appropriate for public walk-up-and-use applications. A walk-up-and-use display must be discoverable, such that first-time users can use the system without any training, flexible, and not fatiguing, especially in the case of longer-term interactions. One mechanism for defining gesture sets for walk-up-and-use interactions is a participatory design method called gesture elicitation. This method has been used to identify several user-generated gesture sets and shown that user-generated sets are preferred by users over those defined by system designers. However, for these studies to be successfully implemented in walk-up-and-use applications, there is a need to understand which components of these gestures are semantically meaningful (i.e. do users distinguish been using their left and right hand, or are those semantically the same thing?). Thus, defining a standardized gesture vocabulary for coding, characterizing, and evaluating gestures is critical. This dissertation presents three gesture elicitation studies for walk-up-and-use displays that employ a novel gesture elicitation methodology, alongside a novel coding scheme for gesture elicitation data that focuses on features most important to users' mental models. Generalizable design principles, based on the three studies, are then derived and presented (e.g. changes in speed are meaningful for scroll actions in walk up and use displays but not for paging or selection). The major

contributions of this work are: (1) an elicitation methodology that aids users in overcoming biases from existing interaction modalities; (2) a better understanding of the gestural features that matter, e.g. that capture the intent of the gestures; and (3) generalizable design principles for walk-up-and-use public displays.

ACKNOWLEDGMENTS

First, I would like to thank the advisors I have had throughout my time at ASU. To Erin Walker, who was there for me and supported me through the last five years, providing guidance, support and encouragement when needed. To Winslow Burleson, who helped me through my comprehensive exams, provided me with the opportunity to intern at MSR, where I started this work, and provided the support and flexibility to pursue my own research interests. To Ellen Campana, who has been tremendously encouraging and has always believed in me, even after she left ASU. To Pat Langley, who provided early mentorship and guidance when I first started this journey over a decade ago.

Second, I would like to thank all of folks at MSR who challenged me and steered me in the right direction: Merrie Ringel Morris, Steven Drucker, Danyel Fisher, and Mary Czerwinski. I also like to thank my committee members for their valuable feedback throughout the process: Kurt VanLehn, Mary Lou Maher, and Stacey Kuznetsov.

Third, I would like to thank all of those at Intel that provided me with the time and the opportunity to work on my dissertation while working full time, especially my manager, Jim Firby, and those who respected the time I set aside to work: Gwen Christian, Wendell Innis, Beth Ann Hockey, and Tope Ogundele.

Finally, I would like to thank my collaborators and friends for all of their support, encouragement and work throughout the years: David Piorkowski, David Lovis-McMahon, Shawn Nikkila, Ryan Spicer and Sean Burdick. A big thank you goes to my family, who encouraged me to finish the journey I started so many years ago.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure | Page

CHAPTER 1

INTRODUCTION

Post-WIMP (windows, icons, menus, pointers) interfaces have existed for over a decade, although in limited use. Beginning in 2006 with the release of the Nintendo Wii, there has been a market explosion of tracking platforms for post-WIMP interaction. As the technology progressed, the time between releases of novel interaction technologies has been decreasing dramatically (*Figure 1*). NielsenWire estimated that as of 2011, 96.7% of US households had at least one TV (Nielsen, 2011) and as of January 2012, Microsoft had sold 18 million Kinect units (Takahashi, 2012). The Kinect gesture-based system began in gaming environments, but free space gesture capabilities have since been incorporated into desktop interaction. For example, Leap Motion and Intel RealSense partnered with laptop manufacturers, such as Asus, HP, and Acer to make gestural interaction available with the purchase of consumer PCs (Acer, 2015; Baldwin, 2013).

More recently, gesture and voice interaction technology has been integrated into drones, virtual reality (VR)/ augmented reality (AR) technologies and robotics. In early 2016, Yuneec launched the Typhoon H drone, which integrated the Intel RealSense hardware and SDK. Also in early 2016, Microsoft and Oculus both released VR products using various gestural input methods. The past few years have been witness to a significant introduction of robotic greeters into public areas such as airports and malls (Evangelista, 2016; Read, 2017). These greeter robots currently use both conversational and touch interfaces as input modalities. Some even have face detection capabilities, although they haven't yet incorporated free space gestural interaction.

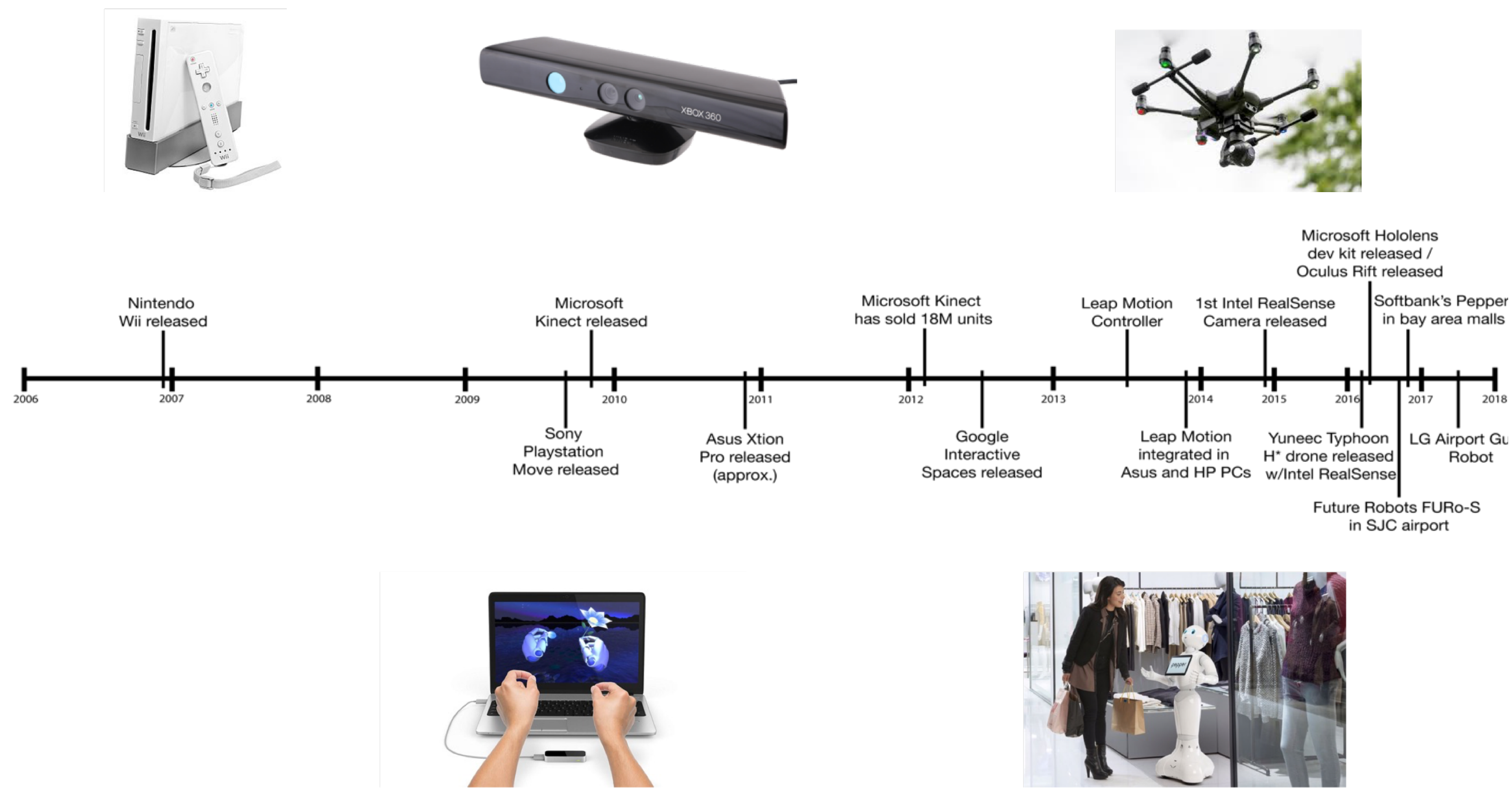


Figure 1. Timeline of key post-WIMP interaction hardware release dates.

Gestural and voice input technologies are increasingly becoming part of daily life, yet the consumer market is still struggling to find compelling use cases and contexts for gesture. One of the reasons for this is that there is still a dearth of generalizable design guidelines that help make gestural input easy-to-use and intuitive. Some research has been conducted that addresses the design issues associated with the demand for increasingly user-centered, gesture-based interactions, but there are still many open questions. To date, designs for new modalities have primarily been built upon previous systems and what little is known about them. While significant research has been conducted to better understand desktop and touch interactions, and results from this research have been incorporated into touchscreen computer interfaces for home and office use, free space gestures have the potential to go well beyond these interfaces. Despite some possible limitations, free space gestures can have rich potential for full body interaction, by allowing for diverse natural user input that people use in human-to-human communication daily. Limitations may include lack of tactile feedback, difficulty in discovering gestures, fatigue, and/or gestural ambiguity (the latter three are addressed later in this document). Therefore, significant gestural interaction research must be conducted to fill critical knowledge gaps relating to each of these issues. We need to better understand how these issues impact whether and how researchers and interface designers can optimize system designs for user-centered everyday gesture-based applications. In tandem with the rise of interface options is an increasingly pressing need for insight into gestural interaction – this need exists across the spectrum from gesture discovery and definition to characterization and recommendation. This dissertation focuses on answering some of these questions, specifically around increasing discoverability, reducing fatigue, and developing a better understanding of users' mental models when it comes to the features of gestures that matter most to their desire to convey meaning.

Participatory design methods (Schuler & Namioka, 1993), specifically gesture elicitation, can be used to appropriately identify gestures that are guessable by a larger user-base. Existing contemporary research in free-space gestures has limited utility for whole body interaction: the bulk of this research is designer-focused, limited by current technologies, and developed primarily for hand gesture. Although there has been some research conducted recently in full-body or foot gestures for limited contexts of use (Alexander, Han, Judd, Irani, & Subramanian, 2012; Grace et al., 2013, 2017). Morris et al., found that researcher-defined gestures may not always match up well to user-preferred gestures (Morris, Wobbrock, & Wilson, 2010). Gesture elicitation studies (Alexander et al., 2012; E, E, Landay, & Cauchard, 2017a; Findlater, Lee, & Wobbrock, 2012; Micire, Desai, Courtemanche, Tsui, & Yanco, 2009; Morris et al., 2010; M. Nielsen, Störring, Moeslund, & Granum, 2004; Wobbrock, Morris, & Wilson, 2009) have successfully combated designer-centered approaches to touch interaction but have not yet been widely used in the context of free-space gestures. These studies have also identified that the methodology may constrain users, through bias or habit (called legacy bias), with existing modalities such as desktop and mouse (Epps, Lichman, & Wu, 2006; Morris, 2012; Morris et al., 2014; Wobbrock et al., 2009). These biases may provide insights into guessable gestures, but may not take full advantage of this novel interaction modality. This suggests ample room, and need, for creative thinking about new possibilities for designing free space gestural interactions for walk-up-and-use interactions. The research community could benefit from user-centered input as demand grows for "intuitive" interfaces across novel modalities and human cultures.

One critical area of research is in developing walk-up-and-use interactions that use free space gestures instead of touch or desktop interfaces. Because gestural walk-up-and-use interactions do not require users to come into direct contact with the system, they are less likely to break or become unreliable with use, although this also means that these systems do not provide tactile feedback to users. Additionally, walk-up-and-use systems in contexts such as museums, malls and airports can provide users with an easy and intuitive means of gathering relevant information (e.g. information about a museum's archives, the location of a store of interest in a shopping center, or which gate your flight is leaving out of). For gestural interactions to be intuitive in a walk-up-and-use context, in which the system must be self-explanatory and not require prior training, several of the limitations of free-space gestural interactions (difficulty in discovering gestures, fatigue, and gestural ambiguity) must be addressed. This means the gestures must be: A) Discoverable or guessable – as defined by Wobbrock et al. (Wobbrock, Aung, Rothrock, & Myers, 2005), a guessable symbol, or in this case gesture is one which allows a user to access the intended referent via that guess despite a lack of prior knowledge of the gesture. By definition, walk-up-and-use displays are self-explanatory, so that anyone can use the system without any prior training, making it possible to interact and look up information quickly. Discoverability is both important and still a challenge with full-body interactions (Cafaro et al., 2014; Cafaro, Lyons, & Antle, 2018; Norman, 2010). B) Easy-to-use – a gesture that is not fatiguing and that is easy for participants to physically accomplish. Limitations of existing technologies often require users to perform fatiguing and awkward gestures that are prone to result in the "gorilla arm" effect (Boring, Jurmu, & Butz, 2009; J. D. Hincapié-Ramos, Guo, & Irani, 2014a; J. Hincapié-Ramos, Guo, Moghadasian, & Irani, 2014; Ruiz & Vogel, 2015). This discourages users to interact with a display for any length of time. Many elicitation studies ask participants to rate gestures based on ease (e.g. (Felberbaum & Lanir, 2018; Wobbrock et al., 2009) C) Flexible and reliable – such that user variability is supported, to account for influences such as culture (Cauchard, E, Zhai, & Landay, 2015; E, E, Landay, & Cauchard, 2017b), resulting in a low number of errors, and high efficiency with little to no gestural ambiguity (U. Oh & Findlater, 2013; Sharp, Keskin, Robertson, Taylor, & Shotton, 2015; Wobbrock et al., 2005). To better design gestures for walk-up-and-use interactions, we must therefore better understand which components of these gestures are semantically meaningful to users and define a standardized vocabulary for coding, characterizing, and evaluating gestures. This will allow us to better train classifiers that focus on the correct gestural features that distinguish gestures in users' mental models. A lack of understanding of users' mental models also means that there are no generalizable design principles, requiring a new gesture elicitation study to be conducted for each new system, making the gesture set contextually useful only for the system it was build.

In this dissertation, I present three studies using a novel gesture elicitation methodology and contribute to the discovery and characterization of a user-centered, guessable gesture set(s) for full body interaction in walk-up-and-use systems. Unlike traditional gesture elicitation methodologies that ask for only one gesture per action from participants, the new methodology asks participants for multiple gestures per action to reduce bias from existing modalities, while still producing easily guessable gestures.

The goal of this approach is to define a standardized gesture vocabulary and qualitative coding scheme that will help identify, confirm, characterize, and offer design support for, genuinely "intuitive" interface design across walk-up-and-use post-WIMP interfaces. This research, therefore, aims to answer the following questions:

RQ 1) How do we modify gesture elicitation to reduce legacy bias?

RQ 2) Which gestural features matter to users and how do they influence a user's mental model about that gesture?

RQ 3) What are the set of design principles that can be used in the future to design gestural interfaces that are discoverable, easy-to-use-and flexible for public displays?

This document is organized as follows: Chapter 2 provides a comprehensive overview of the related work in the design of free space gesture sets, the use of gesture elicitation methodologies and analysis of their results, how fatigue influences gesture performance, and an overview of how human movement is coded and analyzed. Chapter 3 discusses the first of three studies conducted as part of this dissertation work. This study was focused on reducing legacy bias through increasing production and explores the use of priming in gesture elicitation studies. Chapter 4 presents the second elicitation study conducted, which focused on identifying physical fatigue in gesture and how it relates to user preference and discoverability. Chapter 5 presents the third and final study conducted, which was focused on understanding users' mental models in more detail. The primary goal of this final study was to better understand which gestural features are important to users and how these features influence users' mental models, with the aim of better informing data collection efforts used to construct classifiers for gesture recognition. Finally, I close with a meta-discussion of themes that emerged across the three studies and the set of design principles derived (Chapter 6), and a summary of the contributions of this work (Chapter 7).

The primary contributions of this thesis are: 1) a modified gesture elicitation methodology using priming and production to reduce legacy bias, 2) a qualitative coding scheme that better captures users' mental models. 3) a set of generalizable design principles that can be used in the future to design gestural interfaces that are discoverable, easy-to-use and flexible for public displays.

CHAPTER 2

BACKGROUND AND RELATED WORK

In this section, I provide an introductory overview of the work that has been conducted in designing free space gestures for various interactions, both through designer specified approaches, as well as via participatory design through gesture elicitation. Additionally, I will provide an overview of some of the key elements that affect user preference of free space gestures, such as cultural influences and physical fatigue. Finally, I'll provide an overview of how gestures have been coded and analyzed in the past, and why these methods are insufficient for coding full-body gestural interaction aimed at identifying critical gestural features based on users' mental models.

2.1 Free Space Gesture Sets

Some research has already been conducted to develop free hand gestural languages, but many of these studies have been limited in imagination, do not taken full advantage of the new interaction modality (such as constraining users to the use of hands only interaction), require extensive training, or are in other ways inappropriate for walk-up-and-use interactions (e.g. requiring users to lie on their backs during the interaction). For example, Blackshaw et al. developed a gestural language, addressing hand gestures for manipulating an actuated surface (Blackshaw, DeVincenzi, Lakatos, Leithinger, & Ishii, 2011), but their grammar included only five actions: scaling, selection, translation, rotation, and direct manipulation. Such a hands-only grammar has significant limitations for free space modalities: it excludes users who lack use of their hands, and fails to accommodate alternate sources for and types of gesture. Similarly, Zigelbaum et al. worked with the Oblong system (The system also used in the film Minority Report) to develop *g-stalt*, a comprehensive yet relatively simple gestural interface (Zigelbaum, Browning, Leithinger, Bau, & Ishii, 2010) that includes 20 gestures inspired by gesture studies on human-to-human communication. Other studies have developed gestural languages that require extensive training, which are not appropriate for walk-up-and-use interactions. One such system, Charade (Baudel & Beaudouin-lafon, 1993), was developed to allow users to give presentations with hand gestures using a data glove to recognize a complex data set. Unfortunately, in addition to requiring extensive training, Charade is also limited to applications associated with a very specific data set, suggesting associated gestures are neither intuitive nor easily generalizable. In a recent survey article, Johnson-Glenberg provides a set of design principles for VR in education, but only a few are limited to gesture interaction (Johnson-Glenberg, 2018).

Additionally, much of this research has been designer specified instead of developed through participatory design methods, such as gesture elicitation, that are more likely to provide insight into discoverable gestures. In each of these systems designers developed gestural languages without input from users, and the tools required considerable training before use. The limits of such studies (only hands, not intuitive), suggest a potentially significant gap between design and use trends, an increasing liability that adds to the need to identify user expectations for interaction with increasingly ubiquitous, novel interaction systems.

User-centered gesture research has also been conducted, but it is difficult to apply the results to walk-up-and-use interactions since most studies have not focused on this context of use. Oh et al., for instance, looked at whether users prefer using *hover* or *swipe* for different menu layouts in a Hands-Up system with a ceiling projection (J. H. Oh et al., 2012).

They found that it depended upon the projection layout: users preferred *hover* for radial layouts and had no preference for list layouts. While such research does acknowledge the need to understand user preferences, it is still artificially constrained, which limits its utility in walk-up-and-use scenarios (such as public and/or educational displays), since it assumes the user is lying down to interact with a ceiling projection and using an application with fairly limited scope and few menu options.

Another study, conducted by Spiro, uses a webcam-based crowdsourcing application called Motion Chain (Spiro, 2012) to study gestural motion. The system is similar to the ESP game developed by von Ahn, L. and Dabbish (von Ahn & Dabbish, 2004). The application aims to aggregate a large corpus of gestures with potential utility in machine learning, and runs on any computer with a webcam. The application allows users to play one of two games: Charades or Chains. Both are based on the popular children's game, "Telephone." The model has utility and, again, suggests that valuable data comes from users. A similar approach can be used to create a large database to help "teach" machines to classify patterns across variables. The current limitation of the form is that it neither identifies a full gesture set for walk-up-and-use systems, nor distinguishes between what end-users might consider "good" or "bad" gestures.

These examples demonstrate that research does exist that may have relevance in free space gesture sets and post-WIMP interaction design, but that the focus of existing efforts has been too limited, both technologically and methodologically, to be of significant use in everyday contexts. There is increasing need to develop and validate a method for eliciting discoverable, human-preferred gesture sets for new modalities. This need, in tandem with exponentially increasing market demand for user-centered technology, suggests that user expectations—and design preferences—should inform research. Investigating free space gestures requires finding new, less limited elicitation methodologies and gesture set definition(s), and then validating potentially discoverable gestures in ways that mimic potential everyday interaction with ubiquitous systems. Such research could usefully inform design not only in walk-up-and-use displays, but lay a foundation for exploring additional free space applications yet to be discovered.

The goal of this dissertation research, then, is to identify the gestural features that are most important to users and to leverage that understanding to specify at least one gesture set that could support free space interactions in walk-up-and-use applications. The first order of business when exploring novel modalities is to explore them using novel methods, to potentially reduce the impact of legacy bias, which will be discussed further in the next chapter. The nature of this research is, therefore, necessarily inclusive: initial gesture studies are exploratory, both in terms of method and results, and their purpose is to generate possibilities that can be validated later. The long-term aim is to identify a genuinely intuitive human gesture set that is discoverable by a large user base. For this reason, a new gesture elicitation approach must be used.

2.2 Gesture Elicitation

User-generated input is critical for any gesture set approaching ubiquity. Many prior free-space gesture systems (Baudel & Beaudouin-lafon, 1993; Blackshaw et al., 2011; J. H. Oh et al., 2012; Zigelbaum et al., 2010) are designer-specified, despite Morris et al. (Morris et al., 2010) showing that user-suggested gestures are not only different from and simpler than researcher-authored gestures, but are in fact preferred over those gesture sets defined by

HCI professionals. This is a significant finding, underscoring the need for more user-centered input in design assumptions.

In lieu of designer-driven approaches, a participatory design methodology called *gesture elicitation* has been used in several studies to design a variety of gestural and speech interfaces. Gesture elicitation presents users with *referents* (an action's effect) and asks the participant to provide *symbols* (the interactions that could result in that referent). Historically, most gesture elicitation studies have only asked participants to produce one symbol per referent (or one one-handed gesture, and one two-handed gesture per referent).

To identify a non-conflicting gesture set, commonly proposed gestures are aggregated across participants for each referent by calculating an agreement score (Vatavu & Wobbrock, 2015, 2016; Wobbrock et al., 2005). Wobbrock et al. (Wobbrock et al., 2005) proposed the use of the following formula to compute the agreement scores of proposed gestures:

$$\text{Eq. 1: } A_r = \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|}\right)^2$$

In the equation above, *r* is the referent in the set of all referents *R*, P_r is the set of proposed gestures for referent *r*, and P_i is the subset of identical gestures from P_r. This agreement score was proposed for studies in which each participant proposed only one gesture per referent and was further refined by Findlater et al. (Findlater et al., 2012), in which the authors added two correcting factors, which make the equation more accurate as shown in (Vatavu & Wobbrock, 2015). The corrected agreement rate used is the following:

$$\text{Eq. 2: } AR_r = \frac{\sum_{P_i \subseteq P_r} \frac{1}{2}|P_i|(|P_i| - 1)}{\frac{1}{2}|P_r|(|P_r| - 1)}$$

Vatavu and Wobbrock also introduced two additional measures (Vatavu & Wobbrock, 2015), a between group *coagreement rate*($CR_{(r_1,r_2)}$) and *disagreement rate* (DR_r), which are as follows:

$$\text{Eq. 3:} CR_{(r_1,r_2)} = \frac{\sum_{i=1}^{n} \delta_{i,1} \cdot \delta_{i,2}}{n}, n = \frac{1}{2}|P|(|P|-1)$$

$$\text{Eq. 4: } DR_r = -\frac{|P|}{|P|-1} \sum_{P_i} \left(\frac{|P_i|}{|P|}\right)^2 + \frac{|P|}{|P|-1}$$

or more simply

$$DR_r = 1 - AR_r$$

The coagreement rate is used to understand the level of shared agreement between two referents. In the formula for CR, r_1 and $, r_2$ stand for the referents for which the coagreement rate is being computed and n denotes the number of pairs of participants. $\delta_{i,1}$ takes the value of 1 if the i-th pair of participants agree for referent r_1 and 0 otherwise. The same is true for $\delta_{i,2}$. The authors also defined a *k-coagreement rate* for $k > 2$ referents.

Vatavu and Wobbrock further define the *between-group coagreement rate* that can be used to compare groups of participants within a study or across studies (e.g. to compare differences between men and women or across age groups) (Vatavu & Wobbrock, 2016). This formula is as follows:

$$\text{Eq. 5: } CR_b(G_1, G_2, \dots, G_k) = \frac{\sum_{i=1}^{k} \sum_{j=i+1}^{k} \sum_{p=1}^{|G_i|} \sum_{q=1}^{|G_j|} \delta_{p,q}}{\sum_{i=1}^{k} \sum_{j=i+1}^{k} |G_i| \cdot |G_j|}$$

where $G_1, G_2, \dots, G_k$ are the different groups being compared with k $=$ total number of groups being compared, $\delta_{p,q}$ is Kronecker's notation that evaluates to either 0 or

1 depending on whether participants p and q are in agreement or not and the sum goes for all pairs of participants selected from all pairs of groups G_i and G_j, $1 \leq i < j \leq k$.

By going back and analyzing data from previous studies, the authors found that there is a difference between technical and non-technical participants in terms of agreement, which shows that legacy bias does have an effect on the gestures that users specify and their agreement with one another. They also found difference in agreement between men and women and found that touch gestures have higher agreement than free hand gestures, and free hand gestures have more agreement than full body.

To accommodate studies in which an arbitrary number of interactions are proposed per referent, Morris (Morris, 2012) proposed the *max-consensus* and *consensus-distinct* metrics. The max-consensus ratio is equivalent to the percentage of participants that suggest the most popular proposed interaction for a referent and is written as follows:

$$\text{Eq. 6: Max-consensus} = max\left(\forall_{P_i \subseteq P_r}\left(\frac{|P_i|}{|P_r|}\right)\right)$$

The consensus-distinct metric is the percent of the distinct interactions proposed for a given referent (or referent/modality combination) that achieved a specific consensus threshold among participants. Usually, a default consensus threshold of two is used – this means that at least two participants proposed the same interaction. These metrics provide a peak and spread of agreement. These various formulas can be used to identify non-conflicting gesture sets from gesture elicitation studies that users prefer over designer-specified sets.

Gesture elicitation has been used in a wide variety of emerging interaction and sensing technologies, such as tabletop multi-touch gesture sets (Epps, Lichman, & Wu, 2006; Findlater et al., 2012; Micire et al., 2009; M. Nielsen et al., 2004; Wobbrock et al., 2009), physical objects, such as pen-based interactions (Frisch, Heydekorn, & Dachselt, 2009) or mobile phone motion gestures (Ruiz, Li, & Lank, 2011). Few gesture elicitation studies have targeted free space gestures. One, conducted by Alexander et al., used gesture elicitation to identify foot gestures (Alexander et al., 2012); another study used gesture elicitation to identify free-hand gestures for web browsing (Morris, 2012). The latter also introduced a multi-user variant to gesture elicitation to allow multi-person teams to suggest gestures together. To date, there has been little research targeting gestures that could be discoverable in walk-up-and-use public display scenarios.

A couple of exceptions are presented Maher and Lee (Maher & Lee, 2017). In their book, they present two interactive systems, the willful marionette and a walk-up-and-use information display for exploring information on courses and faculty at a university campus. For the walk-up-and-use display, an elicitation study was run in which participants were asked for at least 4 gestures per referent. For the willful marionette, a combination of early prototypes and body-storming were used to generate an initial gesture set, which were then performed by the marionette through a Wizard of Oz approach during a gesture elicitation study with participants. In this study, participants were asked to interact with the marionette spontaneously without being given a task or goal. The approach used for the willful marionette deviates from the way elicitation studies are usually conducted, but regardless provides insight into how users might interact with system requiring gestural input. In both cases, the aim was to identify a gesture set for a particular walk-up-and-use interaction that was easily discoverable.

Eliciting novel gesture sets for new modalities using gesture elicitation also has its challenges. In multi-touch interaction's infancy, for instance, people identified gestures that did not take full advantage of the capabilities of the new modality; they used one finger to interact rather than leveraging multi-touch input (Ryall, Morris, Everitt, Forlines, & Shen, 2006). This suggests researchers must be proactive with their methods and measures and prompt people to think beyond familiar, or technologically constrained, human-computer interactions to combat legacy bias. Cautions are framed by previous research results (Morris, 2012), which found, for instance, that gestures users produced were heavily influenced by mouse interactions. Considering the ubiquity of these technologies, however, this is not surprising—novel systems require novel thinking, and most users resort to habits learned in more familiar scenarios. Users mentioned, for instance, that they would pretend their hand was a mouse they would use to point at objects on screen. Other studies of user-produced gestures have also identified user bias stemming from mouse-based interactions (Epps, Lichman, & Wu, 2006; Wobbrock et al., 2009).

This survey of related work demonstrated that gesture elicitation is a promising methodology to use for identifying a discoverable gesture set for walk-up-and-use interactions, but additional research must be conducted in order to encourage users to take full advantage of the novel interaction and help them avoid generating gestures based on legacy bias.

2.3 Fatigue in Gestures

Although users tend to prefer gestures that are specified by them instead of by designers (Morris et al., 2010), and these gesture sets are usually more memorable and discoverable, prior elicitation studies (Epps, Lichman, & Wu, 2006; Morris, 2012; Morris et al., 2010; Wobbrock et al., 2009) have found that users are biased by familiar interactions (e.g., selecting simple gestures where a finger mimics a mouse), such as WIMP (windows, icons, menus and pointing) interfaces, touch interfaces, or even interactions with existing gestural technologies, such as the Kinect. Legacy bias and can occur because users explicitly want to transfer knowledge from past systems to new ones, which can reduce the physical and mental strain of interacting with a new modality, or because they assume that there are technological limitations that may or may not exist (Morris et al., 2010). For example, due to users' understanding of the technical limitations of 3D motion tracking platforms, they tend to specify large gestures that result in the "gorilla arm" effect or the use of a hover gesture to select items, based on the Kinect and Microsoft 360 gesture set. These gestures, however, violate the primary guidelines provided by Nielsen et al. to avoid fatigue in gestural interfaces (M. Nielsen et al., 2004) and put a significant amount of strain on the shoulder joint, which is known to fatigue faster than the elbow or wrist (Law & Avin, 2010). In general, legacy bias keeps users from producing gesture that take advantage of the new interaction modalities, sensing technologies, and application domains and may propagate existing metaphors that are no longer relevant (similar to how we still use the floppy disk icon to indicate the save action) while also producing gestures that are highly fatiguing and unnatural. For example, "hover" is often one of the first gestures produced in elicitation studies, even though it's often the least preferred gesture and highly fatiguing, as the user is required to hold their hand in the same position for at least 2 seconds (see *Figure 2*).

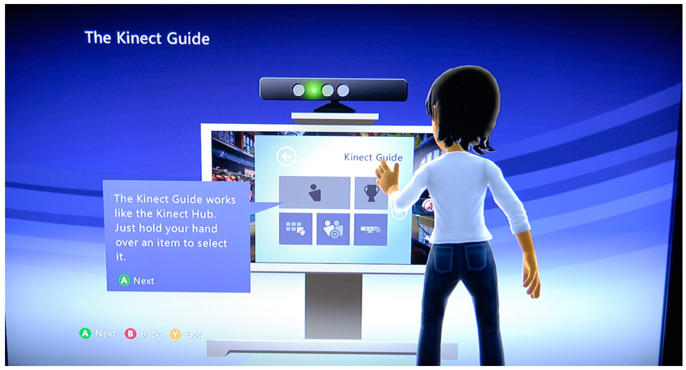


Figure 2. Instructions on how to perform the hover gesture on the Microsoft Kinect.

In an effort to combat legacy bias, Morris introduced the user of *partners* (Morris, 2012; Morris et al., 2014), where two people are asked to produce symbols together so that they can build off of one another's suggestions. The use of partners was inspired by group brainstorming methods used in design as common practice such as in (Dow et al., 2011).

Legacy bias may also have some benefits. Culturally shared metaphors, for example, are one reason for legacy bias and shared metaphors may lead to higher agreement scores (Wobbrock et al., 2009) and higher discoverability and learnability (Wobbrock et al., 2005). Gestures based on metaphors have also been shown to be learned more quickly (Krueger, 1993). However, Cafaro et al. found that legacy bias did not increase discoverability in their study (Cafaro et al., 2018). Köpsel and Bubalo argue that we can benefit from legacy bias, because it can be a helpful tool to gently introduce new forms of interaction, such as gestures or multimodal interfaces, to the general public (Köpsel & Bubalo, 2015). This means that a balance must be struck between the benefits of legacy bias which may lead to more easily discoverable gestures and the drawbacks, which include propagating outdated metaphors and defining gestures that are more fatiguing and less natural for users.

While many ergonomics studies have been conducted for mouse-based interactions, research has only recently shifted its focus on understanding user fatigue in free-space and full body gestural interactions. It is clear, though, that fatigue is an important factor in the adoption of free-space gestural interactions in order to avoid the "gorilla arm" effect ("Gorilla Arm," n.d.). As it has been noted by Lenman et al. (Lenman, Bretzner, & Thuresson, 2002) gestures that require hand or arm movements without support are likely to be difficult for users to repeat or perform for extended periods of time without fatigue. Nielsen et al., who provide an overview of the main principles in ergonomics, also note that interfaces should avoid forcing users to use repetition or stay in a static position (M. Nielsen et al., 2004). They provide the following guidelines for fatigue avoidance in gestural interfaces:

- Avoid outer positions
- Relax muscles.
- Relaxed neutral position is in the middle between outer positions
- Avoid repetition
- Avoid staying in static position
- Avoid internal and external force on joints that may stop body fluids.

Some researchers have begun to apply ergonomics to both touch and free-hand gestures. For example, Hoggan et al. have evaluated various features of pinch gestures to determine which features have the largest effect on ergonomics (Hoggan, Nacenta, et al., 2013). In a related study, Hoggan et al. similarly evaluated rotations performed with the index finger and thumb (Hoggan, Williamson, et al., 2013). In both of these studies, the researchers found that distance, angle, direction and position have significant effects on the ergonomic failure rates and movement speeds of pinch and rotation gestures. Kölsch et al. have evaluated the postural comfort of free-hand gestures at stomach height in the horizontal plane and found that users are more comfortable when they do not have to reach far from their bodies and with interfaces that use both hands for interactions over ones that constrain the user to single-hand interactions (Kölsch, Beall, & Turk, 2003). Cabral et al. evaluated the use of gestures in virtual reality environments and compared the performance of gestures to a mouse (Cabral, Morimoto, & Zuffo, 2005). They found that even the completion time of simple pointing tasks were slower than when using a mouse and that using gestures for a short time could result in fatigue. Existing relevant principles of ergonomics should continue to be applied to free-space and full body gestures. Finally, Lui and Thomas ran two experiments: one with command gestures and another with pointing and selection gestures (X. Liu & Thomas, 2017). In their first experiment, participants were asked to rate how tiring a gesture was and how appealing it was after being given four pairs of gestures, one considered less physically difficult to perform than the other (the four pairs were, with the more difficult gesture second: pinch / grab, finger tap / palm tap, swipe left / swipe up, make a small circle / make a large circle). In the second experiment, participants were asked to select objects randomly from a 3D grid under two conditions – one in which the objects were close together, and another in which they were 2.2x further apart. They found that the more fatiguing a gesture is perceived to be, the less appealing it is to participants.

Qualitative fatigue is often measured through the Borg CR10 scale (Borg, 1998). Additionally, work has been conducted in quantitatively evaluating gesture fatigue for arm movements. Hincapié-Ramos et al introduced a novel measure, the Consumed Endurance

(CE) metric, which is the ratio of the interaction time and the computed endurance time (J. D. Hincapié-Ramos, Guo, Moghadasian, & Irani, 2014). The endurance is the duration that a muscle can sustain a level of contraction before requiring rest. In general, measures of arm fatigue focus on the shoulder, as this is the joint in the arm that's most likely to become fatigued first (Law & Avin, 2010). This is also true of CE. In their study, the authors found that using a bent arm in the vertical center of the body are less fatiguing. Additionally, putting items that are often interacted with in regions lower in the interaction plane and closer to the arm being used will also reduce fatigue. Finally, the authors evaluated a novel keyboard layout, the SEATO, which is less fatiguing than a QWERTY layout for gestural interaction without sacrificing speed and with minimal impact on error rate. This method has been validated against the Borg CR10 scale, but can currently only evaluate arm fatigue. Additionally, the authors have provided a tool that uses input from the Microsoft Kinect camera to calculated the CE metric for different poses (J. D. Hincapié-Ramos, Guo, & Irani, 2014b).

Jang et al. took a different approach (Jang, Stuerzlinger, Ambike, & Ramani, 2017). CE only quantifies fatigue in the moment and can only calculate physical fatigue, not perceived fatigue. Because of this, Jang et al. introduced another metric, the three-compartment muscle (TCM) model, as well as an easier low cost way to estimate max shoulder torque, which is required by the model. To measure max shoulder torque, experimenters measured the time to fatigue by asking users to hold a weight directly out from their body horizontally for as long as they can. The assumption is that users exert constant shoulder torque for the duration of the exercise. They validated this measurement by comparing it to the measurement from a dynamometer. The authors then conducted an experiment with a mid-air pointing task aimed at collecting subjective fatigue ratings using the Borg CR10 scale and biomechanical measurements of fatigue to generate a cumulative fatigue model that describes the relationship between the two. Rest periods (between-subject) and vertical location of interaction (within-subject) were factors that were tested in this experiment. They also compared their measure to CE, showing that CE was less accurate than their method, as it underestimated fatigue levels by not accounting for cumulative fatigue, with estimates dropping to nearly 0 during rest periods. They also found that, while their model was accurate, it was highly impacted by the interaction zone (vertical location of interaction).

To understand the effect of fatigue in elicitation studies, Ruiz and Vogel asked participants to wear wrist weights while performing gestures during an elicitation study (Ruiz & Vogel, 2015). They found that wearing wrist weights deterred users from performing large arm gestures and instead elicited a larger variety of gestures (e.g. using feet to indicate selections, or head nods) and gestures with lower CE scores than when participants did not wear the weights. However, gestures had low agreement scores and it's unclear whether wearing weights while performing a gesture is a valid substitute for prolonged use.

To reduce fatigue from free-hand gestural interactions with public displays, researchers have started to explore at-your-side gestures (M. Liu, Nancel, & Vogel, 2015; Siddhpuria, Katsuragawa, Wallace, & Lank, 2017). In (M. Liu et al., 2015), Liu, Nancel and Vogel present Gunslinger, which uses leap motion cameras mounted on the thighs of the user to track subtle at-your-side gestures, and explore the use of Gunslinger in tandem with touch gestures. In this study, the gesture set for both touch and free-space gestures are defined by the researchers and hand gestures are limited to finger and hand poses. More recently Siddhpuria et al. conducted an elicitation study for at-your-side gestures with the intention of using a smart watch as the tracking input (Siddhpuria et al., 2017). In this case, gestures were limited to motion gestures, and could not suggest poses, as these wouldn't be recognized. Although participants generally rated their gestures as highly natural, having low fatigue and as generally being socially acceptable, the authors found a low level of consensus for elicited gestures.

2.4 Describing and Annotating Human Movement

Various fields of research have sought to develop diverse classifications and taxonomies of human movement to help shed light on the features of gesture that are important to people when communicating with one another or expressing themselves. Understanding the features that are important in human-to-computer communication is just as important, as this understanding can help designers gain insight into users' mental models and the gestural features that semantically distinguish gestures from one another. This will allow designers to better create gesture sets that are discoverable, not fatiguing and flexible and this information can be used to more accurately train gesture recognition classifiers. A better understanding of users' mental models through a clear and consistent annotation scheme, will also allow for the creation of generalizable design principles. These design principles can be used as a starting point for new systems in various contexts, instead of having to run a new elicitation study for each system that is designed and built.

When trying to understand gesture as it relates to speech, various researchers, such as Efron (Efron, 1941), Kendon (Kendon, 1986), and McNeill and Levy (McNeill, 1992; McNeill & Levy, 1982), have come up with different taxonomies of gestures. In McNeill and Levy's taxonomy, gestures are classified as either:

- **Metaphoric**: Gestures that portray the speaker's idea, while relating only indirectly to the content of the interaction.
- **Symbolic:** Standardized gestures, complete within themselves, without the need to accompany speech, and may be less likely to transfer across cultural boundaries.
- **Iconic**: Represent the content of speech, such as an entity or action – e.g. making a gesture of bending back an object in space while saying "and he bent it back."
- **Deictic**: Gestures which locates references and objects with respect to a spatial reference frame, e.g. "pointing."
- **Beats**: Rhythmic hand gestures accompanying speech that can signal to a listener which parts of the speech are more important than others.

Deixis and iconicity are two dimensions of gesture meaning, not really different gestures (Lascarides & Stone, 2009; McNeill, 2005) and their meanings are largely improvised in context (Lascarides & Stone, 2009). Unlike the communicative gestures presented so far, *manipulative gestures* are gestures whose purpose is to control some object or entity through a tight coupling of the movement of the hand and object being manipulated (Pavlovic, Sharma, & Huang, 1997). See (Karam & Schraefel, 2005) for a more comprehensive overview of the different gesture taxonomies.

While these taxonomies may provide insight into whether gestures will be easily identified across cultural boundaries, using just this taxonomy does not provide the level of detail necessary to understand relevant features of the gestures that are both important to users and can be used to train classifiers for future systems.

3][4][5][6][7] [8] [9]

00:16:46:28 # [[/ so<oo> **you're in th**e kitchen] [['n' then there's as<sss>* /] [/ the **back stairc***]]

1 2a 2b

1. iconic; RH sl.spread B PTD moves AB from RC to hold directly above knees; <moving into the kitchen; then, loc. in kitchen>
2a. deictic vector§ - *aborted*;; RH G takes off from end point of #1; points and moves RtoL across C; head tilts to look left in the same direction the hand is moving; <move from kitchen area to bottom of back staircase>
2b. deictic vector§ - *aborted*;; RH G takes off from end point of #2a; in one smooth motion: points to R while moving TB, curves to point & move up; <move up back staircase>.

[10] [11][12

00:16:50:12 [oh I **forgot to say**]

emblem-ish; A-hand moves to sternum & holds; <(woops)>

12][13]

00:16:51:10 [[when you come **through** the* / / #]

iconic$^{\Omega}$ - *aborted*;; BH/mirror 5-hands PTB 5-hands start together PTB @chest, move AB and apart to a more PTC/PTB position on either side of CC; <open & move through interior double doors>

[14][15] [16][17][18] 19] [20][21][23]
22

00:16:2:29 [[when yo**u ente**r the ^house ^from the ^front /] [an**nd you**<ou> / **open**n the / doors / with t]] [he*

1 2

1. iconic; BH/mirror 5-hands start together PTB @chest, move AB but *do not move apart*; 'superimposed' beats, including a big upward & forward head movement at the end on "front"; <enter the house, but not stop short of going through the interior double doors>
2. iconic - *repeat & complete* $^{\Omega}$; BH/mirror 5-hands PTB 5-hands start together PTB @chest, move AB to a **mid-stroke hold**, then move apart to PAB on either side of C @far L&R; <open & move through interior double doors>

Figure 3. Sample of a psycholinguistic analysis transcript from (Quek et al., 2002).

Unlike most studies that look at and code for only hand gestures, Quek et al. (Quek et al., 2002) and McNeill (McNeill, 1992), in their research, define gestures to include not only hand movements, but also full body movements, facial expressions and gaze shifts. To analyze gestures used in human-to-human communication, Quek et al. annotated and analyzed their data as is seen in *Figure 3* and *Figure 4*. While their work codes the 2D location of the gesture in their video, and classifies the gesture based on taxonomy, it does not provide the level of detail for classifying 3D gestures. It's also heavily tied to the linguistic description the users are providing, as opposed to being a primary interaction modality.

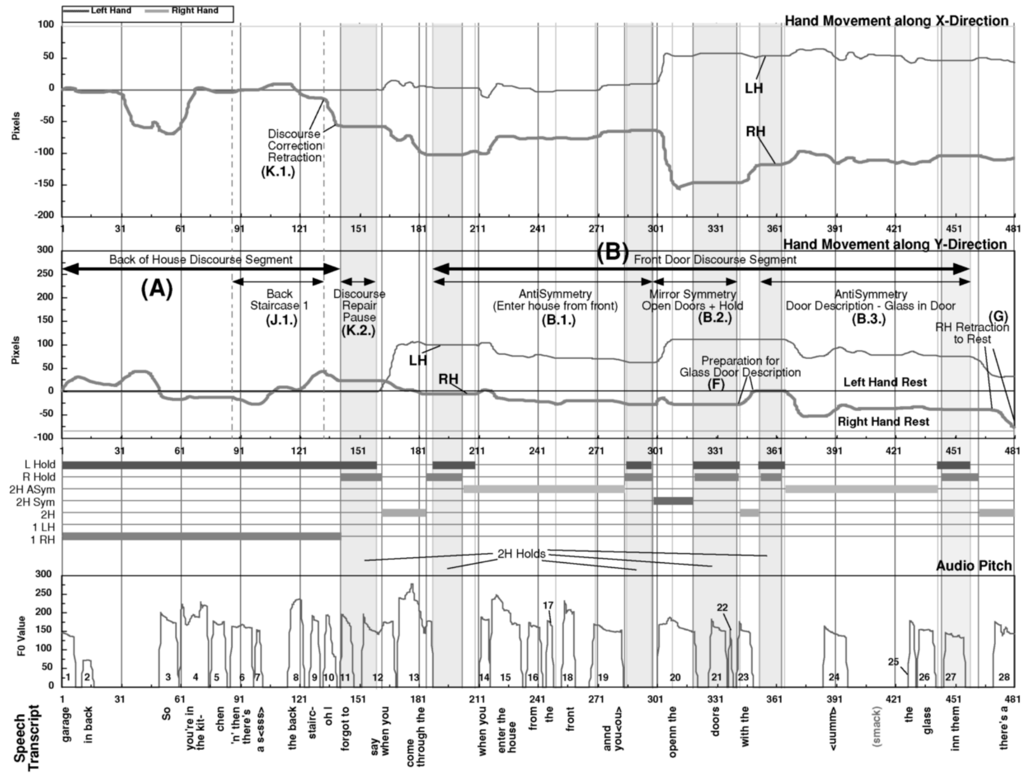


Figure 4. Hand position, handedness analysis, and F_00 graphs for several frames of a psychoanalysis transcript from (Quek et al., 2002).

In dance, one of the primary ways of expressing gesture is through Laban movement analysis (LMA) (Laban & Ullmann, 1971). The major categories used in Laban notation are *body*, *effort*, *shape*, and *space*.

- **Body:** Describes how the movement flows through the body.
- **Effort**: is about the dynamics of the movement, and describes things such as flow (free flowing or bound), weight (light or heavy), time (sustained or quick), space (direct or indirect movement).

- **Shape:** is about the form, and the shape qualities include things such as rising / sinking (vertical direction), spreading / enclosing (horizontal direction), advancing / retreating (sagittal direction).
- **Space**: describes the body's motion in connection to the environment and concerns itself with the kinesphere, spatial intention, and geometric observations of the movement.

Several works in HCI have used LMA to better understand gestures (Loke, Larssen, Robertson, & Edwards, 2007; Moen, 2005). In (Morris et al., 2014), the authors found that the content on the screen may influence the time and weight of gestures. Additionally, complementary referents (e.g. scrolling up and scrolling down), will affect the shape (e.g. participants may recommend a rising vs a shrinking gesture). While I draw inspiration from LMA, especially around gesture path and flow, it is difficult to discriminate some of the features of LMA, such as speed (Grijincu, Nacenta, & Kristensson, 2014) and Laban features must be calibrated to each individual user (Sikora & Burleson, 2017).

Recently, Mcaweeney et al. have sought to study ways in which visual representations of gestures can be used to communicate gestures supported by a system to users (McAweeney, Zhang, & Nebeling, 2018). They began by reviewing 30 gesture elicitation studies for touch, air and tangible interaction to produce a taxonomy of gestures. From this review, they classified gestures into six dimensions (Body context, Environmental context, Perspective, Frame, Color, and Gesture elements) and 26 categories. They then conducted an elicitation study to understand user mental models. In this study, they asked participants to draw out several representations of gestures based on videos of the acted-out gestures and then show this representation to a partner, who acted it out. Once the partner correctly acted out the gesture, the pair co-designed a representation of the gesture for each referent. The authors used the taxonomy from study one to code the final 150 co-designed representations. Analyzing the content of the representations, they came up with five features that nearly all representations articulated: Time, Position, Posture, Motion, and Touch. This study provides good insights into features that matter, but focused on touch gestures and on communicating those gestures to users via diagrams, not in the features that matter when training classifiers to discriminate between user gestures.

Others, such as Krupka et al. have proposed their own simple languages to define hand gestures and have developed a set of tools to show the language can be used for both development and gesture recognition (Krupka et al., 2017). Krupka et al.'s language is based on 4 predicates - pointing direction, relative location, fingertip touching and finger flexion - used to describe hand poses at six points on the hand, the fingertips, and the palm center. A gesture is defined by a sequence of hand poses (see *Figure 5* below for an example). The authors claim that their language can express the basic signs of American Sign Language (ASL) phonology and the basis poses used in several current commercial systems. The pipeline was trained on more than 360,000 annotated images of hand poses. Their recognition algorithm results in 96% detection accuracy after some training, making this system impractical for walk-up-and-use systems, since users first need practice and training with the system. Their system is also only focused on hand gestures, meaning that the language would need to be extended for full body gestures.

```xml
<Gesture Name="RotateGesture">
    <Gesture.Segments>
        <IdleGestureSegment Name="Idle"/>
        <SingleHandPose Name="LockObjectPose">
            <FingersPose Context="Thumb, Index" Flexion="Open" Direction="Forward"/>
            <FingersLocalityRelation Context="Index" LocalityRelation="Above" OtherContext="Thumb"/>
            <FingersTangencyRelation Context="Index" TangencyRelation="NotTouching" OtherContext="Thumb"/>
        </SingleHandPose>
        <SingleHandPose Name="RotateObjectPose">
            <FingersPose Context="Thumb, Index" Flexion="Open" Direction="Forward"/>
            <FingersLocalityRelation Context="Index" LocalityRelation="Right" OtherContext="Thumb"/>
            <FingersTangencyRelation Context="Index" TangencyRelation="NotTouching" OtherContext="Thumb"/>
        </SingleHandPose>
    </Gesture.Segments>

    <Gesture.SegmentsConnections>
        <SegmentConnections From="Idle" To="Idle, LockObject"/>
        <SegmentConnections From="LockObject" To="RotateObject"/>
        <SegmentConnections From="RotateObject" To="Idle"/>
    </Gesture.SegmentsConnections>
</Gesture>
```

Figure 5. Example of a coded hand gestures for (Krupka et al., 2017).

Unlike (Krupka et al., 2017), Hachaj and Ogiela present a Gesture Description Language (GDL) for full body gestures and a classifier for the GDL that was evaluated on 1,600 movements with an 80.5-98.5% rate of accuracy (Hachaj & Ogiela, 2014). The GDL uses key frame descriptions to classify gestures (See *Figure 6* below for example gestures specified using the GDL). The GDL they describe is not readily designer friendly and due to its very specific nature, could accidentally reject gestures that users believe are similar or identical based on their mental models. It provides no support for designers to evaluate which features matter and which do not (e.g. a swipe with the hand would be encoded differently than a swipe with the arm, even though they might be the same to the user). Their system was evaluated on an arbitrary set of gestures that most systems out there don't currently use (e.g. hand on head, rotating clockwise and counter-clockwise) and a few that are used in existing walk-up-and-use systems, such as clapping and waving. None of the gestures they evaluated had leg movements, even though the GDL is supposed to cover full body motion.

```
RULE RightElbow.x[0] > Torso.x[0] & RightHand.x[0] > Torso.x[0]
& RightHand.y[0] > RightElbow.y[0] & abs(RightHand.x[0] - RightElbow.x[0]) < 50
& abs(RightShoulder.y[0] - RightElbow.y[0]) < 50 THEN RightHandPsi
RULE LeftElbow.x[0] < Torso.x[0] & LeftHand.x[0] < Torso.x[0]
& LeftHand.y[0] > LeftElbow.y[0] & abs(LeftHand.x[0] - LeftElbow.x[0]) < 50
& abs(LeftShoulder.y[0] - LeftElbow.y[0]) < 50 THEN LeftHandPsi
RULE RightHandPsi & LeftHandPsi THEN Psi
```

```
RULE distance(RightHand.xyz[0], LeftHand.xyz[0]) < 100 THEN HandsTogether
RULE distance(RightHand.xyz[0], LeftHand.xyz[0]) >= 100 THEN HandsSeparate
RULE sequenceexists("[HandsSeparate,0.5][HandsTogether,0.5][HandsSeparate,0.5]") THEN Clapping
```

Figure 6. Examples from the rule-based GDL (Hachaj & Ogiela, 2014). The second example shows the description for recognizing a hand clap.

Finally, researchers studying gesture-based interfaces have focused on identifying gestures that are learnable, discoverable, immediately usable (Long, Landay, & Rowe, 1999), and memorable (Nacenta, Kamber, Qiang, & Kristensson, 2013). Grijincu et al. have therefore focused on enabling further analysis of memorability through the annotation, sharing and descriptive analysis of the set of gestures that were generated by participants in previous work by the authors (Grijincu et al., 2014).

While all of this work aims to understand and classify gestures, many require a large data set to train on, and few can be used in designing full body interactions for public displays by understanding users' mental models (e.g. of which features are most important to them) to group gestures accordingly and better train classifiers. Instead, we need to identify a gesture taxonomy that allows for full body motion and can provide better insight into which features matter most so as to not end up with trained classifiers that make distinctions between gestures that users do not make.

CHAPTER 3

STUDY 1: ADDRESSING LEGACY BIAS

As previously mentioned, *legacy bias* is the influence that existing modalities such as desktop and mouse interactions, or even existing free-space and full body technologies, have on the gestures that people suggest for novel interactions (Morris et al., 2014). This study contributed to the identification of legacy bias and explored ways in which researchers can conduct elicitation studies to reduce or eliminate legacy bias through changes in priming and production, which addresses RQ 1 presented in the introduction of this document. Additionally, this study explored the relationship between legacy bias and discoverability and ease-of-use.

In this chapter, I present a modified elicitation methodology based on (Wobbrock et al., 2005, 2009) to encourage creativity from participants via *priming* and *production* with the aim of reducing legacy bias in gesture elicitation studies and identifying new gestures appropriate for depth-camera-based interactions. The primary goal is to encourage participants to come up with gestures that don't suffer from the drawbacks of legacy bias (specifically the elicitation of fatiguing and unnatural gestures), while still allowing participants to come up with gestures that will be easily discoverable. The user of *priming* was inspired by North et al.'s finding that users who performed a task with physical objects before using a touch-table were less likely to use only pointing-based interactions (North, Dwyer, Lee, Fisher, & Isenberg, 2009). Priming is already a well-established phenomenon, in which prior stimuli influence how people perceive and react to future stimuli (Wentura & Degner, 2010). In this study, I explore the use of both kinesthetic and video priming, as described in the modifications section below (3.1.1).

For production, standard gesture elicitation methods ask users to produce a single gesture to cause the action depicted by the referent prompt (Wobbrock et al., 2005, 2009). In a few cases, participants have been asked to perform a single gesture with one hand, and another gesture with two hands. To reduce legacy bias, in this study I modified the *production* methodology such that each participant was asked to produce as many gestures as they could come up with (with the experimenter aiming for between five and nine gestures each). By eliciting multiple gestures per referent, the elicitation method attempted to reduce bias from other interaction modalities, and encourage user creativity, similar to how increased production is used in ideation processes in design. The hypothesis is that increasing production would lead to identifying gestures that are more appropriate for free space interactions while maintaining discoverability, and improving reliability and ease of use by getting participants to move beyond the first few gestures, which are more likely to be influenced by legacy bias. Iteration and increased production have already been shown to be beneficial in design (Dow et al., 2010, 2011; Dow, Heddleston, & Klemmer, 2009; J. Nielsen, 1993; Terry & Mynatt, 2002). In the ESP image-labeling game, "taboo" words are listed to prevent users from always proposing obvious tags for images (von Ahn & Dabbish, 2004). Additionally, Dow et al. found that forcing designers to generate a large amount of initial ideas resulted in final designs that were better than if the designers only generated a small number of initial ideas (Dow et al., 2011). Therefore, requiring multiple suggestions for each referent would help participants move beyond biases stemming from years of mouse-based interactions and allow them to move past the first few biased gestures to produce more novel gestures or iterate on initial gestures, ultimately specifying more appropriate gestures for the interaction modality.

The primary aim of this study is to address RQ 1 (How do we modify gesture elicitation to reduce legacy bias?). A secondary aim was to understand what variables affect the types of gestures produced (RQ 2 and RQ 3). Therefore, the following research questions were addressed in this study:

1) How does increased gesture *production* affect user preference and the types of gestures elicited?
2) Does *priming* affect the number or types of gestures elicited?
3) Which variables (such as the layout of the objects on screen and the number and type of objects users are asked to interact with) affect the gestures produced?

3.1 Method

To reduce legacy bias, we modified the standard elicitation methodology (Wobbrock et al., 2005, 2009) with *priming* and *production*, as mentioned previously. In the study, we chose to focus on a single possible walk-up-and-use application that would allow us to further explore question 3 and would benefit from user-generated input: *faceted browsing*. Faceted browsing (Yee, Swearingen, Li, & Hearst, 2003) is a method for navigation that allows users to explore large collections of data by filtering along multiple categories and is a common way of interacting with product catalogues on e-commerce sites and other metadata-rich "big data" collections.

Participants filled out a short demographic survey before they came to the lab. Once the participants came to the lab, they were provided with a brief introduction of what they would do that day, followed by either the video priming, or kinesthetic priming for those in one of the two priming conditions. The remainder of the study was spent on the gesture elicitation, with the first referent used as a warm up so that participants got used to the task and to thinking aloud. Participants were videotaped by several cameras mounted in the ceiling for the duration of the study. The study lasted a total of 90 minutes and participants were compensated with a $50 gift card to the company store where the study was conducted.

Figure 7. Still image from the video used in the video

3.1.1 Modifications for Eliciting Creativity

Priming: Participants were randomly assigned to one of three conditions: *no priming*, *video priming*, or *video and kinesthetic priming*. Six participants were assigned to each priming condition, and five assigned to the no priming condition. Participants in the *no priming* condition were given a brief amount of time (about 5 minutes) to relax and become accustomed to the space. *Video priming* participants were asked to watch a three-and-a-half-minute video comprised of clips of gestures in different settings, including gestures used in various sports, comedy acts and silent movies, and landing signals for aircraft (*Figure 7*). *Video and kinesthetic priming* participants were first shown the same video as in the video priming condition and then asked to perform the following 15 gestures:

- Walk in a circle. (whole body, free-space, standing)
- Tug your left ear with your right hand. (self-contact)
- Hop (standing, whole body

- Sit down on the couch.
- Make an X on the table with your right hand.
- Stand up.
- Act like a plane.
- Act like an elephant.
- Turn in place.
- March.
- Raise your right leg.
- Raise your left leg.
- Pound your chest like a gorilla.
- Walk to the screen.
- Throw a pillow at the wall.

Production: In this study, we asked participants to produce as many gestures as they could come up with for each referent. After each gesture suggestion, participants were prompted to try to produce another gesture until participants produced at least five gestures per referent and ran out of ideas. Repetition of gestures was allowed between referents, but not within the same referent. Once participants had produced the required number of minimum gestures, they were permitted to stop if they claimed to be out of ideas. After each participant was finished, they were asked to reflect upon their suggestions and indicate a favorite and least favorite gesture. Notably, and unlike Wobbrock et al. (Wobbrock et al., 2009), participants were not asked to immediately rate each proposed gesture. In pilot studies, we asked participants to rate each gesture on a seven point Likert Scale for the following:

- How well do you think the gesture matched the action performed in the video?
- How easy was the gesture to make?
- How fatiguing do you think this gesture would be if performed repeatedly?
- How creative do you think this gesture is?

This was found to impede their creative flow, which conflicts with the goal of creative elicitation, and was removed from the final study.

3.1.2 Participants

17 participants were recruited from outside of the organization in which the study was conducted at a major tech company in the Seattle area. None had any prior experience with depth camera systems (e.g. Microsoft Kinect). Of those 59% were female, and ages ranged from 18 – 48 years (mean = 30).

3.1.3 Environment

The study took place in a laboratory setting arranged like a living room to make users feel more comfortable and encourage creativity. The room had a 63" wall-mounted TV + Kinect opposite a couch (see *Figure 8*). They were told that they could use the space available to them, as well as any objects within the space. Participants were video recorded from multiple cameras mounted to the ceiling to ensure that all gestures were visible by at least one camera without any occlusion.

3.1.4 Referents

Participants were randomly assigned to one of four video referent orders. Each of the four video orders were randomly generated to account for any order effects. All participants were shown all 14 videos, unless the participant ran out of time (only two of the 17 participants ran over on time, with participants taking between 45 and 90 minutes to get through all referents). Participants were instructed that gestures were defined as any sort of movement (i.e., not necessarily limited to hands/arms), and asked to think aloud during the study. They were encouraged to ignore existing technological limitations -- to pretend that any computer they were interacting with could perceive and understand any sort of human movement they proposed.

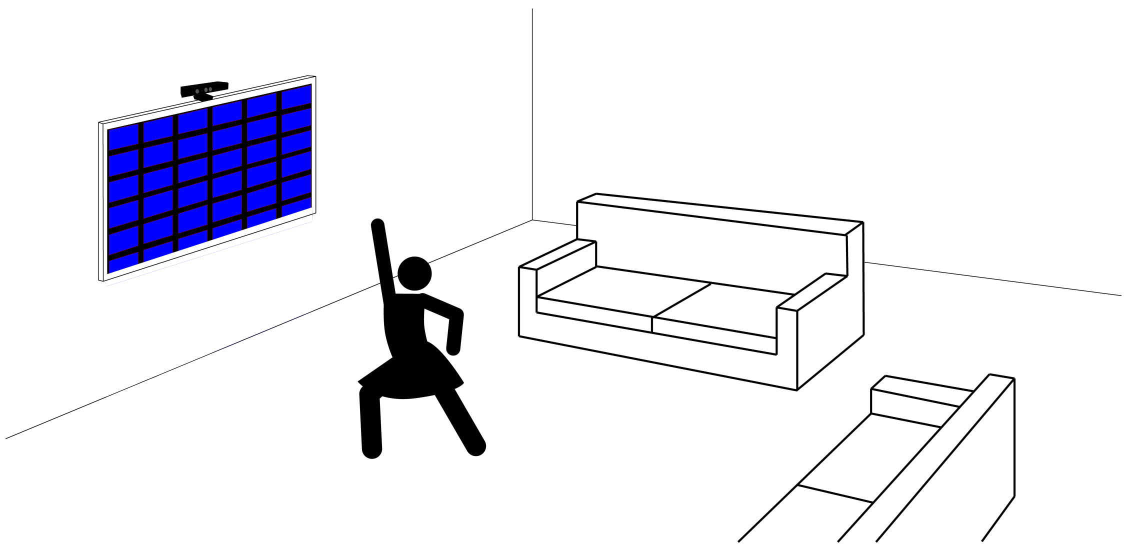


Figure 8. Environment in which the gesture elicitation study took place. Silhouette by Moriah Rich, from The Noun Project.

In order to assess how different variables affected the gestures produced by participants, the layout of the objects, the type of objects (text or abstract shapes), and the number of objects were varied across referents. The study asked every participant to look at four layout styles: linear, grid, radial and a three-column), with each participant seeing all four layout types. See *Figure 9* for an example of each of the linear, radial and three-column layouts that were shown to participants and *Figure 10* for an example of the grid layout for the selection action. To probe whether different data types might necessitate different gestures, we looked for differences between how people treated *text* and *abstract* objects. To test whether the number of objects on screen made a difference in participants' gestural choices, we looked at whether participants' behaviors varied by layout and/or navigation requirements.

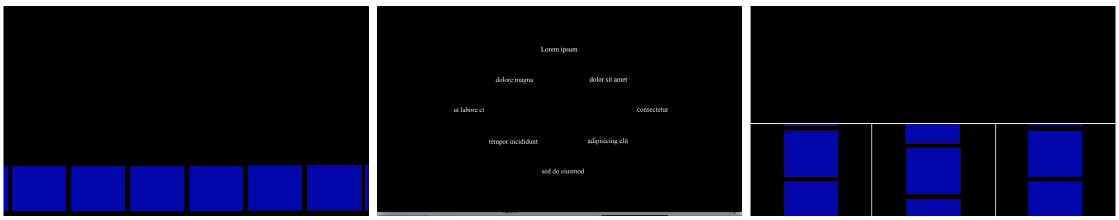


Figure 9. Three examples of referents which vary by layout, number, and type of objects.

Left: Paging action, abstract shapes, linear layout. Center: Selection action, text, radial layout. Right: Scrolling action, abstract shapes, 3-column layout.

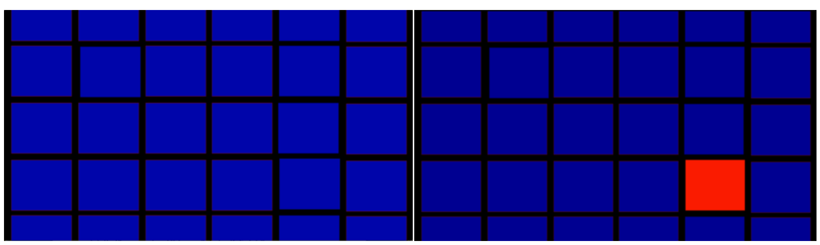


Figure 10. An example of the referent which asked participants to select the square turning red amongst blue squares.

3.1.5 Elicitation Process

Participants would see the left image in *Figure 10* and then hear "*Pretend you are selecting one object out of many from a grid. Here is an example"* after which the image would switch to the right image in *Figure 10*. There was no targeted interaction with the video during the time the participant was suggesting actions; the video was simply placed on repeat. The participant would then be asked to act *as many gestures as they came up with* that could cause the selection to occur. One participant, for instance, produced the following gestures, in order: (1) hover

over the object to select it, (2) grab at the object and pull back, (3) pull object towards a "hot spot" on the screen, (4) point at the object with the right arm, (5) point with both arms at the object.

3.2 Coding and Analysis

During the study, each gesture was coded by the observer as the participant was performing the gestures (see Appendix A for an example of the form used by the observing experimenter). The observer made notes on the position of the user in the space (location related to the screen, orientation, whether the person was sitting or standing), the gesture primitive, the type of motion, the type of gesture, the hand configuration, the body parts used, the side of the body that was used, the palm position, whether the gesture mapped to the body or a specific location in the room, and for those referents that had a speed component, whether speed was based on movement or position. This coding was then used to help with gesture segmentation and to speed up the qualitatively coding after the study. Whether the person was sitting or standing was added after piloting, as we noticed that some users would describe the gesture they would perform instead of acting it out, even after being encouraged to act it out.

We began by coding a short description of each gesture, such as "point toe" or "jab at it", but as we were doing so, we realized that this would result in too many unique gestures would ignore gestures that were semantically similar to participants. we then considered Laban Movement Analysis (LMA), and how instructors describe movement in martial arts and climbing instruction as two of the researcher had backgrounds as instructors. However, LMA requires significant training to use, and as mentioned in the related work, discriminating some of the features of LMA, such as speed, is difficult (Grijincu et al., 2014).

In martial arts and climbing, the nuances of the mechanical movement matter, whereas in everyday gestural interaction people aren't aware of the subtleties of their movement. Instead, we chose to use elements of LMA in coding for speed, direction of motion, and coding for relative or absolute movement, but simplified the process. Research in touch interactions tells us that hand configuration might matter, although usually not the number of fingers used (Morris et al., 2010; Wobbrock et al., 2009), so we also coded for this feature to determine if there are differences between free-space and touch interaction. Finally, codes were normalized across researchers (e.g. speak vs. voice, or hover vs. hold). If a gesture did not fit into existing codes for a feature, a new one was created and the list of possible codes was built up in this way.

The final features that were encoded are listed in Table 1. Whether the person was sitting or standing was noted in a separate "notes" column in cases where the user was sitting instead of acting out the gesture, since this case was relatively rare. The direction of motion was normalized to a list of possible values, as shown in the example for easier analysis. Whether the gesture mapped to the user's body or to the room was encoded in the final coding scheme with the gesture being labeled as either relative or absolute. An example of a relative gesture is swiping anywhere in the space as long as they were swiping up. An absolute gesture was one where, for example, the user must be standing in front of a column or object to perform the gesture.

A subset of the gestures was pairwise coded with an additional researcher to ensure consistency. This data was then used to identify the median position of the favorite and least favorite gestures, and the most popular and least popular gestures. Any comments that participants made during the think-aloud process that highlighted the reasons for either defining the gestures or preferring them, or any metaphors used were transcribed and recorded as notes during the coding process. Metaphors were especially important to extract, as gestures based off of metaphors have been shown to be learned more quickly (Krueger, 1993). Quotes from the transcripts were extracted and grouped in themes that emerged (e.g. if several participants mentioned the same metaphors or motivation for defining a gesture).

Table 1

List of features that were coded in Study 1 for each gesture with example values

Feature	Examples
Gesture Primitive	Point, swipe, kick
Gesture # in Video	1, 2, 3, 7
Gesture Type	Repeating, Sequence, Simultaneous
Body Part Used	Arm, hand, leg, head, full body
Gesture Direction	To an object, forward, back, right, left
Side of body	Right, left, both
Hand Config.	Fist, flat hand, 1-finger point
Palm Position	In, out, up, down
Speed Mapping	To position, to movement
Relative / Absolute?	Relative, absolute
Favorite / Least favorite	Empty, favorite, least favorite

3.3 Results

In this section, we present the results as they relate to the research questions laid out at the beginning of the chapter. Section 3.3.1 presents results related to increased production and Section 3.3.2 presents results related to priming (RQ 1). Section 3.3.3 presents qualitative observations and quotes from participants based on emerging themes. From these, some features that do and don't matter for users' mental models become clear (RQ 2). Additionally, several design principles begin to emerge (RQ 3).

Table 2

Percentage of gesture primitives for each part of the body for Study 1

Body part used	% of primitives using body part
Arms (including hands, elbows, forearms, fingers)	60.80%
Full body	22.92%
Legs (including feet, toes, knees)	7.49%
Head	4.20%
Voice	3.09%
Eyes	1.51%

3.3.1 Effect of Increased Production

To evaluate the effect of increased production in gesture elicitation, we began by looking at the variety of gestures produced by participants, in order to evaluate whether they moved beyond gestures that were solely inspired by existing interactions. Even though all participants defined *pointing* for selection and *swipe* for paging and scrolling, participants also proposed gestures using hands, arms, legs and the head, or whole-body movement (in the space, or leaning, turning, and twisting). 35% of gestures specified did not use arms, for example. See Table 2 for a breakdown of participant's proposals by part of the body used. Additionally, a total of 133 unique gesture primitives were identified during the qualitative coding, although many were only suggested by one participant each, such as "doggy paddle", "lunge", "fishing", and flying like an "airplane". Only 37.6% (n = 50) of gesture primitives were mentioned by two or more participants.

The study also confirmed the value of changing the production method to ask participants to define a large number of gestures by looking at the average position of favorite and least favorite gestures across *all* participants. The median number of gestures produced per participant was 7 (SD=1.35) and the median position of the favorite gesture was 3 (SD = 2.33), indicating that users' first suggestion generally was not their "best." Even though the production of more than one symbol has value, we noticed that some participants' gestures diminished in variety: Having pointed with their hand up, for example, they might then point with their hand down, and then point again with a fist. This suggests there is room for improvement in the production method.

3.3.2 Effect of priming

For priming, we conducted a one-way ANOVA between subjects to test the effect of priming on the quantity of gestures produced by participants. Priming was found to have no effect on the quantity of gestures produced (F (2, 14) = 1.17, p = .339). See Table 3 for the means and standard deviations of the quantity of gestures produced across participants per condition. However, priming did seem to impact users' gesture suggestions. For example, only users who saw the video priming produced gestures where their two arms were rolled about each other – a motion highly similar to one in a priming video clip; non-primed users did not produce this gesture. Without further analysis, we cannot conclude that this transference was beneficial and did not introduce unwanted biases.

Table 3

Means and standard deviations for each priming condition for the number of gestures produced across participants

Condition	Mean # Gestures Produced	Standard Deviation
Kinesthetic Priming	6.95	1.36
Video Priming	7.09	1.46
No Priming (Control)	6.34	1.07

Additionally, participants who were kinesthetically primed tended to produce more gestures that involved moving about the room than others, although this trend fell short of statistical significance (F (2, 14) = 0.51, p = .612). See Table 4 for means and standard deviations for each. Replication with more participants is necessary to verify the impact of priming on the types of gestures produced.

Table 4

Means and standard deviations for each priming condition for gestures involving moving about the room

Condition	Mean # Gestures for move	Standard Deviation
Kinesthetic Priming	5.83	7.25
Video Priming	3.33	5.41
No Priming (Control)	4.20	5.06

3.3.3 Users' Mental Models and Emerging Design Principles

This section summarizes our qualitative findings regarding users' mental models about faceted browsing (RQ2), and emerging trends in their gesture suggestions that could be used as design principles in the future (RQ3). Specifically, we look at how different variables, such as the layout of the objects on screen, and the different types and number of objects, affect the gestures participants produce. We also highlight unexpected findings in free-space and full-bodied interactions that emerged. This section is broken down into the high level themes that emerged during analysis. As previously mentioned, portions of the users' speech were transcribed during the coding process if that speech contained information around the motivations for defining a gesture, contained insights into the features that mattered to users, or contained metaphors that users drew upon when defining the gestures. These quotes were grouped into themes based on common features of gestures produced or motivations for defining those gesture. At least three participants (20%) had to mention the same feature or motivation for it to be counted as a separate theme. A total of six themes emerged in this way:

- Similarities Across Selecting, Scrolling and Paging
- Gestures on the Body, Gestures in Space
- Gestures for Size and Speed
- Gestures for Parallelizing Data Exploration
- Objects vs. Text
- Concerns about Ambiguity

Below, we go into detail for each theme.

Theme 1: Similarities Across Selecting, Scrolling and Paging

Although hover was a common first gesture produced for selection tasks, it was rarely selected as a favorite gesture. Unsurprisingly, for selecting items, nearly all participants chose a pointing gesture. For scrolling and paging through items, all participants proposed a swipe gesture. The most common gesture for all three of these actions were similar to gestures found in touch interactions. For swiping, the direction of motion was more important than the hand used or the palm position. We found that although participants generally had their palm facing the direction of motion towards the beginning of the study, they tended to lose track of this as the study progressed. Participants largely avoided gestures that caused them to face away from the screen for scrolling, but not for paging. For example, spinning in a circle was not used for scrolling, but it was used for paging. This seems to be due to the fact that users see scrolling as a more continuous action for which they need more control, unlike paging, which is a discrete action (the expectation was that the screen would change one page at a time).

Theme 2: Gestures on the Body, Gestures in Space

While most participants stayed rooted in place and used gestured that largely involved waving their hands in the air, some participants began to find other options. One clever choice was body mapping. Some participants (P6, P7) mapped the length of their forearm to the items on screen and began to tap parts of their bodies to select individual items: the wrist would be the lowest item; the elbow was the top. Another participant came up with a method of choosing one of three columns by touching his left or right shoulder, or in between the two shoulders. Such gestures truly take advantage of (and push the boundaries of) the capabilities of depth-camera input.

Several participants began to disambiguate multiple columns (e.g. the right most image in *Figure 9*) by moving in space: a step to the left indicated a motion in the left column, while a step right indicated that they wished to manipulate the right.

Theme 3: Gestures for Size and Speed

Several of the participants mentioned that less precise gestures could be used when objects were larger (such as selecting columns, or with the objects instead of the text), or when there were fewer objects. P3 stated, "if there were a lot of items we felt like we would have to scroll through or line up with the ones we wanted instead of just being able to pick it directly." Conversely, a few of the participants made bigger gestures when there were more objects. P4 and P6 noted that in a grid layout, they might want to sometimes move a single row, or the entire grid. A "big" swipe would grab all the rows and columns; while a smaller swipe might only take a single row or an individual item. Other participants used two-handed gestures when there were more items. P12, for example, stated that he wanted to push the content down with both hands because there was a lot of content there, and that a one-handed swipe would be preferred when there were a smaller number of items.

Two-handed gestures may have meant "big" to some users, but others felt they meant "fast." P9 and P12 both used two-handed swipes to signify that they wanted the system to scroll faster than when swiping with one hand. Many other participants mapped either the speed or the angle of their gesture to the speed of the scroll. For example, the higher one lifts their arm above the point where it's parallel to the ground, the faster the scroll.

Theme 4: Gestures for Parallelizing Data Exploration

We had initially suspected that participants would use their dominant hand to make gestures (76% were right-handed). We were startled to see participants freely switching between hands—indeed, sometimes they would do a gesture, then repeat it with the opposite hand. About a third of the participants preferentially used the hand closest to where the task started regardless of hand-dominance. For example, P12, who was right handed, said "I'm using my left hand since it's sort of on the left side. It feels more natural than pointing with the right arm."

Several participants (P3, P4, P6, P7) indicated that they wanted to be able to multi-task and manipulate multiple objects in different columns at once. To allow for this, many participants defined single-handed gestures that relied on relative body positioning to identify the column being used. For example, swiping with the left hand on the left side of the body, while also swiping with the right hand on the right side of the body allowed participants to simultaneously manipulate both the right and left columns. Either hand swiping directly in front of the body could manipulate the middle column; or, if users wanted to manipulate all three columns at once, a foot was sometimes slid forward or back for the middle column.

Theme 5: Objects vs. Text

We had wondered whether participants would manipulate abstract objects differently than they would text, since both data types abound in faceted browsing tasks. P6, said that moving text didn't "feel right," but that moving objects seemed intuitive. P7, who came up with a way to map scrolling up and down to his right forearm, said that he liked this scrolling technique for objects, but would not use it for text; he preferred to pull the words towards the direction he wanted them to go. Some participants (P3, P8, P16) noted that because the text was smaller than the objects, they felt like they needed finer grained or more precise motions for selecting text. P8 also stated "you can scroll more aggressively with items rather than text just because of the size."

Theme 6: Concerns about Ambiguity

Several of our participants were concerned that their gestures might be seen as ambiguous by the system. P8 said that two-handed gestures felt clearer to her than one-handed, noting that with one-handed gestures, "… you get this wave back and forth going and it's not clear which way you're going…" Two-handed gestures were considered more deliberate.

Participants were particularly concerned that non-hand/arm-based gestures would not be interpreted with adequate precision. For example, many participants suggested that they might be able to nod their heads at an item to select it. P2 identified the appeal behind this small movement, "the less I have to do the better," but conceded that it might be hard to specify *which* item he was selecting in this manner. Similarly, several participants thought of kicking towards the screen for selection. They, too, worried that the system might have trouble figuring out precisely which point they were aiming toward.

3.4 Discussion

Looking back at the original research question around priming and production, which was the focus of this study, the following answers emerged:

For production:

Q: How does increased gesture *production* affect user preference and the types of gestures elicited?

A: Users on average preferred their third gesture specified, not the first or second that are more likely to be impacted by legacy bias. Additional, more creative gestures (e.g. not swipe and point) came later in the gesture elicitation sequence for each referent.

For priming:

Q: Does *priming* affect the types of gestures elicited?

A: There is some evidence that priming affects the gestures produced. First off, kinesthetic priming showed a non-statistically significant increase in participants suggesting movement around the space. Second, participants that watched the video or performed the kinesthetic gestures before the study were more likely to suggest similar gestures during the elicitation study (a different kind of bias than the legacy bias from existing interaction modalities).

3.4.1 Effect of Production

In this study we saw ample evidence for the use of production to increase creativity and help combat legacy bias. As was presented in the results, users suggested a wide variety of non-arm gestures and over 100 unique gesture primitive were identified. This variety suggested that asking participants to define a large number of gestures for each prompt mitigated the effect of typical desktop and touch interaction biases (habituated behaviors/thinking) identified in previous gesture elicitation studies (Epps, Lichman, & Wu, 2006; Morris, 2012; Morris et al., 2010; Wobbrock et al., 2009) and exemplifies participants' willingness to propose gestures that take fuller advantage of a free space medium, although hand and arm movements are still most prevalent.

However, even though the production of more than one symbol has value, as mentioned, we also noticed that variety diminished over time. Across referents, participants began to repeat gestures suggested for previous prompts. This could be due participant boredom or fatigue as the study progressed due to the time required by the production stage, which was on average 40 – 60 minutes. It could also be due to the lack of variety across referents, or the level of abstraction of the referents. The finding that users' favorite gesture was typically the third gesture they produced suggests that reducing the number of prompted commands or introducing more variety in an elicitation study might be a better choice for fatigue control than simply reducing the minimum production requirement. The use of specific, rather than generic, data and scenarios might also alleviate subject boredom.

Within a particular referent, participants also suggested multiple gestures with modifications in only one feature (e.g. swiping up with the full arm vs swiping up with only the forearm or hand). These small changes could actually be due to participants refining their gestures, rather than suggesting completely new gestures, data which was not available with the methodology used in this study. Potential modifications for the elicitation methodology are discussed in section 3.4.4 below.

To better understand the effect of the modified production methodology explored in this study, additional research is needed to determine (1) whether the gestures were iteratively refined, (2) whether participants continued to produce unique gestures beyond the first few ideas and (3) whether the most preferred, or favorite gestures may indicate gestures that are easier-to-use and more reliable. These questions are addressed in subsequent studies.

3.4.2 Effect of Priming

While we found that increased production was a promising avenue to pursue for future gesture elicitation studies, the use of priming is less certain. Understanding the impact of specific types of priming on user-elicitation procedures is a rich area for further investigation, and others have started to more closely study it, although such research falls outside the scope of this dissertation.

In research conducted in response to this study, Hoff et al. conducted a study with 30 participants that looked at the effect of kinesthetic priming and production on the symbols provided by individuals (Hoff, Hornecker, Bertel, Weimar, & Media, 2016). For the priming task, they asked participants to clear the room of boxes and to stretch to help the experimenter determine where in the frame of the camera the participant was standing. They found that there was only a small effect of priming on the elicited gestures and concluded that individual variability outweighs the effect of priming. Specifically, the number of legacy gestures produced was not statistically significant across primed and non-pried participants, but the primed participants produced gestures faster and with lower variation. However, their study also found a medium effect size, which a G*power test indicating that a study with 170 participants would need to be conducted to find a statistically significant result. They also asked participants to produce three gestures per referent and they found no effect of production on the resultant gesture set, finding that participants preferred their first gesture specified. Their study may have introduced unintended bias by asking users in different conditions to fill out the demographic survey at different times -- the survey was filled out on a computer and the non-primed group did this before the study, while the primed group after. This is a potentially confounding factor that was introduced in the study, and may have caused the non-primed group to suffer from legacy bias given the recency of using a desktop interaction immediately before the elicitation.

Cafaro et al. used priming with framed guessability, an elicitation methodology that uses embodied allegories to create “frames” or scenarios during the elicitation phase (Cafaro et al., 2018). The aim is to ground the elicitation, such that gestures that are performed by participants are interconnected, increasing discoverability and user preference. To do so, they conducted an experiment with three conditions: funhouse (n = 29), gym (n = 31) and control (n = 29). For both the funhouse and gym conditions, participants were given a scenario (e.g. "being in front of a distorted mirror while visiting a funhouse") and then were primed in the following three ways: 1) Visual Priming: pictures of the scenario were displayed on a screen when they entered the lab. 2) Written Task: Participants were then asked to write five things they would do in that scenario. 3) Embodied Priming: Participants were finally asked to act out what they wrote in the written task. After priming (for those that had priming), participants went through a traditional elicitation methodology. Gesture sets created in each of the conditions was then evaluated for discoverability in situ in a museum setting. Similar to our results, this study found differences in body vs. arm movements based on condition, specifically the gym condition produced a statistically significantly higher number of body movements and lower arm movements. Cafaro et al. also found a significant effect on the gestures produced, in which participants mentioned the scenario in their elicitation process when defining gestures. This is consistent with our study, in which participants who were video primed produced gestures that were similar during elicitation. Finally, they found that gesture sets produced via framed guessability with priming were more discoverable than those produced via the traditional elicitation methodology. Additionally, many of the gestures produced in the control condition were likely a result of legacy bias, where participants mentioned interactions with existing technologies, whereas in the other conditions there were no such mentions. In this study, however, it is not clear which of the priming methodologies had an effect, or whether all three are required for the effect to be observed.

Given the differences in the studies conducted, the inconsistency of the outcomes between the studies and additional confounding factors, more research would need to be conducted to understand the use of priming and how it affects elicitation outcomes. This research is left for future work and outside the scope of this dissertation.

3.4.3. Users' Mental Models and Emerging Design Principles

In the results, we presented six themes that emerged from transcripts of users' comments during the think-aloud process. Each of these themes highlights a feature (RQ2) or design principle (RQ3) that can be leveraged for future exploration and design of walk-up-and-use systems.

Across the 6 themes, the following insights about users' mental models emerged(RQ2):

- **Palm direction** was less important than direction of motion for swipe gestures. This was especially true as swiping gestures became more fluid as they were repeated over and over again. This finding highlights problems with current gesture recognition approaches such as that by Krupka et al., in which palm and finger directions are key components of defining a gesture (Krupka et al., 2017). (Theme 1: Similarities Across Selecting, Scrolling and Paging)
- **Hand-dominance** does not have an effect on the side of the body with which participants choose to perform a gesture. Layout and position on screen are more likely to affect their choice of whether to use a limb on the right or left side of their body for selection tasks. Nearly all users, regardless of hand-dominance preferred using the right hand in tasks that were in the center of the screen or for paging and scrolling tasks. However, hand-dominance may affect speed, which we did not

analyze in this study. In their study, on efficiency of translations and rotations on touch interactions, Nguyen and Kipp found that right-handed movements were much faster. However, the majority of their participants were right-handed, and they hypothesize that this might be why (Nguyen & Kipp, 2014). Similarly, right-handed participants were found to be faster than left-handed participants when it comes to mouse and pointer interaction (Mouloua, Mouloua, Mcconnell, & Hancock, 2018). This theme will be revisited in future studies. (Theme 4: Gestures for Parallelizing Data Exploration)

The following design principles also emerged (RQ3):

- It was common for participants to **map the screen to their own body or to the floor** in front of them. This was likely due to the fact that the screen they were interacting with was large and mapping it to a smaller space, or a space they felt they had better access to made it easier to interact with. While we were surprised by users mapping the screen to the space originally, this may have been one way for users to cope with the "gorilla arm" effect and from keeping their arms outspread for prolonged periods of time, since this is fatiguing for users (Lenman et al., 2002). Mapping interactions to a user's own body allows them to keep much of the interaction in the optimal space for arm movement (Kölsch et al., 2003). (Theme 2: Gestures on the Body, Gestures in Space)
- Participants expected **less precise gestures to work when objects were larger, or when there were fewer objects**. A larger gesture (or a gesture with both hands instead of one) was performed when participants wanted to move more or larger objects on screen or when they wanted to move them faster. This type of interaction

is consistent with the theories of embodied cognition (Lakoff & Johnson, 1980) and embodied interaction (Dourish, 2001). Users leveraged knowledge of real-world physical interactions in which heavier or more objects require larger movements, and translated that same behavior to digital interactions, (Theme 3: Gestures for Size and Speed)

- **Manipulating text was seen as different from manipulating objects** and different gestures were specified for each type of object, even if the action itself was the same (e.g. scrolling through text vs objects). While there are studies that specifically look at using mid-air gestures for text entry (Jones, Alexander, Andreou, Irani, & Subramanian, 2010; Ni, Bowman, & North, 2011), there are no known studies that compare text to abstract objects. As noted in the results, some of the differences observed in the manipulation of text vs. objects was due to difference in size between the two, however many others were due to real-world metaphors (e.g. underlining text or circling a word of interest on a physical page.) (Theme 5: Objects vs Text).
- Many participants expressed concerns about **ambiguity** of gestures and how these could be interpreted by other people or technology (Theme 6). These concerns suggest two things: 1) that current-generation depth-sensing technology may not yet be of adequate resolution for "big data" tasks like faceted browsing. and that 2) Gestures that may be perceived as ambiguous by other people may need additional ways for the system to resolve ambiguity, either through a continuous feedback loop or the use of other modalities, such as speech. Human communication is inherently multimodal, and therefore plenty of research has already been conducted in

multimodal interaction (Cabral et al., 2005; Claude, Cerma, & Carbonell, 1993; Morris, 2012; Quek et al., 2002; Robbe-Reiter, Carbonell, & Dauchy, 2000; Robbe, 1998). User concerns around ambiguity only highlight the continuing need for the design of multimodal systems.

Finally, participants regularly tailored their gestures to the level of specificity of the prompts (i.e., the size and number of generic objects/text items shown). This suggests that elicitation studies couched in data sets of a specific type, size, and quantity might be beneficial for producing task-specific gestures, but does not allow us to create generalizable design principles to help speed up the design process or to create universal gesture systems that can be used across applications on a particular platform. In order to design for platform specific, instead of application specific gesture sets, more research must be conducted into when type, size and quantity of items on screen affect gesture production and when they do not. This study can serve as a start to that line of inquiry by identifying differences between text and objects, a large vs. small number of items on screen, and the perception of fast vs. slow scrolling. Some of these differences will be explored further in Study 3, presented in Chapter 5.

3.4.4 Challenges with Coding and Methodological Modifications

Coding and analyzing participants' gestures produced by this novel methodology proved challenging. During pilot testing, interrupting participants during the production phase interfered with their creative flow; in subsequent sessions, they were neither asked to rate gesture quality on a per-gesture basis, nor interrupted to ask clarifying questions. This resulted in a significantly reduced amount of data that the experimenters were able to use to segment gestures and extract features.

The lack of per-gesture ratings and clarifications reduced available data enough that it was difficult to recommend a final faceted browsing gesture set. The large number of unique gesture primitives identified is both promising and problematic. While it shows improvements in participant's creative output and ability to move beyond legacy bias, it also highlights how difficult it is to group similar gestures together and derive consensus, both across qualitative coders and participants themselves. As was mentioned in the results section, many of the gesture primitives were mentioned by only one participant, with only 37.6% of gestures being mentioned by two or more participants.

Another problem occurred with the fact that breakpoints separating individual gestures were not always clear, as even participants noted that in some cases gestures blended together (e.g. P8's comment that "… you get this wave back and forth going and it's not clear which way you're going."). This made it difficult to distinguish individual movements at any level of granularity above a high level "swipe" and therefore, during the coding process answering questions about the start and end positions of gestures and whether they were meaningful was more difficult.

Finally, as noted previously, there was less variability across gestures even within a referent as participants continued to elicit gestures. Participants during the think-aloud process didn't often talk about the differences between gestures that they were eliciting, so it was impossible to determine whether a gesture was merely a refinement improving gesture a previously elicited gesture, or an entirely new gesture. Lack of clarifying interruptions also impeded the understanding of users' mental models about which aspects of a gesture were or were not integral (e.g., was the use of the left hand rather than the right meaningful, or simply the fact that the hand was swiping?).

To reduce the ambiguity of gesture breakpoints and to provide more insights into the features that matter to users, subsequent studies (presented in future chapters) asked participants to watch video of themselves acting out the gesture suggestions, and then to rate them. In Study 3, presented in Chapter 5, participants are also directly asked if a gesture is a refinement of a previous gesture. Participants were also encouraged to provide more detail during the think-aloud process that can be used to guide the experimenters when conducting qualitative coding of the gestures.

3.5 Summary

The primary contributions of this study are:

1) An exploration into the use of *priming* for gesture elicitation studies.
2) Modifications to the *production* of symbols in gesture elicitation studies to increase creative output.
3) A better understanding of *legacy bias* and how priming and production affect it.

In addition, in this study, a few features were identified as either being important or not to users (hand-dominance, palm direction) and several design principles emerged around how to treat large vs. small objects, text. vs images/objects and mapping a large display to a different or smaller surface to make it more accessible.

The methodological changes to production increased variety, while the addition of priming showed no clear benefit. Additionally, users' preference for gestures that were not their first produced indicate that this methodology may be more likely to identify gestures appropriate for certain walk-up-and use interactions that are expected to go beyond a few seconds or minutes. However, the methodology needs to increase referent variety, so participants do not become bored. Additional changes to the methodology are needed to shed light on users' mental models of gestures without breaking their creative flow during production, such as introducing a video retrospective of the elicitation step, in which participants review their produced gestures, to the study design. These changes are implemented in a subsequent study presented in chapter 5.

CHAPTER 4

STUDY 2: A CLOSER LOOK AT FATIGUE

In the first study presented in Chapter 3, we looked at the effects of priming and production on legacy bias and whether using priming and changes in production improved the gestures produced in elicitation studies of walk-up-and-use interactions. The study showed a positive effect of increasing production and highlighted a few design principles that can be explored further in subsequent studies. While increasing production is a promising avenue for addressing legacy bias, there were still many gestures produced by participants that were affected by legacy bias. For example, many of the gestures performed in Study 1 were large gestures influenced by misunderstandings of the technology or gestures that were carried over from other modalities, such as touch interactions. These gestures, while they might be more discoverable, are also gestures that are prone to the "gorilla arm" effect, and may not be preferred for walk-up-and-use interaction. Additionally, while users were asked for their favorite and least favorite gestures as a way to identify preference, they had many definitions of preference and very little insight was provided on how preference relates to discoverability and fatigue. Therefore, in this study, we focus our attention on examining how users' *preferred gestures* intersected with the *most discoverable gestures*, and how fatigue influences which gestures are preferred by users. The aim of the study was to address RQ 3 (What are the set of design principles that can be used in the future to design gestural interfaces that are discoverable, easy-to-use-and flexible for public displays?) by deconstructing it into the following research questions:

1) **Discoverability**: Is the first gesture defined the one that users prefer on their initial ranking?

2) **Fatigue**: Do users' rankings of preferred gestures change after they have been asked to repeat the gestures for an extended period of time?
3) How do discoverability and fatigue relate to one another?

4.1 Method

To explore the relationship between physiological fatigue and user preference, we conducted a gesture elicitation study with 15 participants (10 females, 5 males). Participants over the age of 18 were recruited from the larger university population to participate in a one hour-long study. All participants were right handed and had various levels of comfort with technology. About two-thirds of them had previously used some kind of free-space gesture recognition system before (such as the Microsoft Kinect or Leap Motion). Participants were asked to wear an Affectiva Q sensor (Picard, n.d.) on their right wrist to measure galvanic skin response (GSR) and provide a quantitative assessment of physical fatigue.

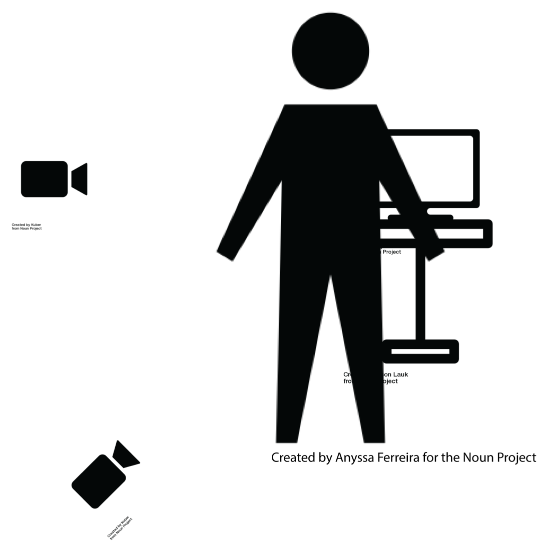


Figure 11. Fatigue study setup. The monitor was 21" and placed on a 3' tall table. The participants were videotaped from the back and diagonally from the front and side.

The study was conducted in a lab. Participants were asked to interact with a 21" monitor that was set up on a three-foot tall table. The surrounding area was large enough for the participant to move around (it was approximately a 6' x 6' open area). Participants were videotaped by two cameras, one set up to videotape from behind the participant, and a second in front of and to the side of the participant (see *Figure 11*). All participants were compensated for their time with a 20-dollar gift card.

At the beginning of the session, participants were asked to complete a demographic survey (See Appendix B for the survey). Then, they were instructed to pretend that they were in a furniture store that had a gesture-based walk-up-and-use display. The display provided customers with the store's inventory, including furniture that was not out on the showroom. They were asked to imagine that the display could pick up any gestures they performed, and to not constrain themselves to current technological limitations. As in previous elicitation studies, participants were asked to think-aloud.

Keeping in line with the positive changes to production from the previous study, each participant was shown six referents, or tasks, and asked to identify at least four symbols for each referent. The tasks were: activate the screen, select an object, scroll, page, undo, and return to the main menu. See *Figure 12* for an example of the selection referent after the item had been selected. The referent to activate the screen was used as demo task to familiarize participants with the protocol of the study. The remaining five tasks were shown to the participants in randomized order to account for any order effects on their ratings. In line with current elicitation study methodologies, no real-time feedback was given to participants' gestures as they were being performed.

Figure 12. Screenshot of the "select" video referent after an item had been selected.

After each task, participants were asked to rank the gestures in ascending order based on preference. They were asked to provide motivation for this order and asked if they felt like they were physically fatigued during the task. Finally, they were asked to imagine doing the gestures for an extended period of time and identify those gestures that would be most fatiguing. See Appendix C for the User Survey. After the survey, the participants were asked to perform each gesture repeatedly for 10 seconds each, explicitly violating the HCI heuristic for repetition (one of Nielsen's main principles of ergonomics (M. Nielsen et al., 2004)). Repeating the gestures for 10 seconds would emulate instances in which the user was, for example, searching through a larger catalogue and would be scrolling for some time. This will likely result in the "gorilla arm" effect for some gestures, in which the user's arm starts to become sore, cramped, etc. from being held up in front of their body for too long and the muscles having to contract for an extended period of time. Once they had repeated all of the gestures for 10 seconds, they were asked again to answer the same survey questionnaire so that their original answers and their new ones could be compared to note any changes in users' preferences.

Table 5

List of features that were coded in Study 2 for each gesture with example values.

Feature	Examples
Gesture Primitive	Point, swipe, kick
Gesture # in Video	1, 2, 3, 7
Primitive Sequence # (for cases in which many primitives were performed to achieve an action).	1,2, 3
Body Part Used	Arm, hand, leg, head, full body
Direction / Path	Up, Down, In, Circular, ZigZag
Hand Configuration	1-Finger, 2-Finger, Fist, Flat Hand
Palm Position	In, out, up, down
Duration of Gesture	The time it took to produce the first gesture.
Duration of Repeated Gestures	Approximate time users repeated the gestures, since repeating for exactly 10 seconds is difficult.
Gesture Type	Deictic, Iconic, Symbolic, etc. (this was not used for the analysis related to this dissertation).

4.2 Coding and Analysis

Gestures produced by participants were segmented and then coded for the features listed in Table 5. In this study, the aim was to better understand discoverability and fatigue and the relationship between them, not to deep dive into users' mental models about the gestures they perform. Therefore, the list of features that were coded was kept to the minimum necessary to distinguish between gestures, while still providing enough information to group similar gestures together through a more consistent set of gesture primitives. The list of possible primitives was built up as new gestures were coded that did not fit into existing primitives. We began with a short list of common gesture primitives that was informed by the most common gestures elicited in Study 1 (e.g. swipe, point, voice commands, etc.) Researched coded all gestures and added codes as needed. This data was then used to calculate the gestural consensus across participants and the diversity of gestures produced.

Additionally, the survey responses before and after repetition were added to the analysis. In this study, participants were not explicitly asked for their favorite or least favorite. Instead, the favorite gesture was assumed to be the first gesture listed in the ranking of the all gestures produced for a referent. Comparing the ranking before and after repetition allowed us to identify any changes in user preference before and after inducing fatigue. The GSR data provided by the Q sensor was not properly baselined for each participant and was, therefore, discarded from analysis.

4.3 Results

In this section, we discuss the results of the qualitative coding of the gestures and the results of the questionnaires for each task in this study. Overall, participants defined a total of 438 gestures, resulting in 499 gesture primitives across all 6 referents. Some gestures had multiple primitives associated with them. For example, "point to an object and then grab it" was coded as one gesture with two gesture primitives: point, grab. Similar to the prior study, all participants defining between 3-7 symbols per referent (with an average number of 5 symbols per referent across participants). During the coding, a total of 72 unique gesture primitives were identified.

4.3.1 Discoverability of Gestures

To better understand which gestures are discoverable, we calculated the agreement score across participants per referent. The agreement score identifies gestures that are most common among all participants, which makes them more likely to be discovered since a large number of participants suggested the gestures. Because participants were asked to identify a variable number of actions for each referent, though, the agreement score proposed by Wobbrock et al. cannot be used (Wobbrock et al., 2009). Instead, we calculate the max-consensus and consensus-distinct scores as introduced by Morris, for each referent (Morris, 2012).

The max-consensus score highlights the most agreed upon gestures (i.e. the gesture that is specified by the largest number of participants for each referent). The consensus-distinct score gives an indication of the variety of gestures specified by participants for a particular referent. A larger consensus-distinct score suggests fewer total unique gestures specified for a particular referent. A referent with both a high max-consensus and a high consensus-distinct score, therefore, is indicative of strong agreement on one particular gesture with few other contending gestures. Table 6 shows the results of both the max-consensus and consensus-distinct scores for each referent, calculated with a consensus threshold of 2 (meaning that at least two participants agreed on a gesture for the task).

In Table 6, both the *activate screen* and *undo* referents have the same gesture primitive ("wave") with the highest max-consensus. Similarly, the *return to main menu* and *scroll* referents both have "slide" as the referent with the highest max-consensus and *page* has "swipe" as the referent with the highest max-consensus, with the primary difference between slide and swipe is that a swipe is performed faster than a slide. This highlights a potential problem with gesture elicitation studies: since each of the referents is designed for independently, there is nothing prohibiting a gesture to be elicited with the highest consensus for multiple prompts.

Table 6

Max-consensus and consensus-distinct scores for each referent in Study 2

Referent	Gesture with Highest Consensus	Max-Consensus	Consensus-distinct
Activate screen	Wave	60%	0.429
Page	Swipe	80%	0.438
Return to main menu	Slide	40%	0.421
Scroll	Slide	73%	0.538
Select	Point	100 %	0.500
Undo	Wave	67%	0.556

The *select* referent had the highest max-consensus score at 100%, meaning that every single participant specified “point” as a gesture primitive for that task, whereas *return to main menu* had a relatively low max-consensus score at 40%, and also the lowest consensus-distinct score for all referents. So while *select* may be heavily influenced by legacy bias, *return to main menu* is not a common gesture interaction used in other modalities, and therefore has higher variability and is less influenced by legacy bias than other actions.

Table 7

Top 3 most elicited gestures for each referent in Study 2 and their consensus scores.

Referent	Gesture	Max-Consensus
Activate screen	Wave	60%
	Point	40%
	Tap	40%
Page	Swipe	80%
	Slide	40%
	Wave	33%
Return to main menu	Slide	40%
	Tap	33%
	Point	27%
Scroll	Slide	73%
	Swipe	67%
	Point	60%
Select	Point	100%
	Tap	40%
	Side	47%
Undo	Wave	67%
	Form X	60%
	Shake	40%

Surprisingly, the consensus-distinct scores do not differ greatly across referents, even in cases where a high max-consensus score is achieved for a particular gesture. That means that the variety of gestures produced for each of the six referents is similar. On average, there were 14 distinct gestures elicited per referent (*SD* = 1.33) by participants with a threshold of 2 and 29 distinct gestures (*SD* = 4.46) if no threshold is applied. To better understand the kinds of gestures produced for each referent, and to help create a unique gesture set for this task given the lack of uniqueness in the most common gestures for each referent, we looked at the top three gestures produced for each task and their consensus scores, which can be seen in Table 7.

By looking at the second and third most common, and therefore, discoverable gesture produced within referents, we can now start to identify a unique gesture set across referents. For example, for *undo* "wave" and "form x" both have similar max-consensus scores, but "form x" does not conflict with *activate screen*. Similarly, "tap" and "slide" for *return to main menu* are similarly discoverable due to similar max-consensus scores, but tap does not conflict with *scroll* or *page*.

4.3.2 Effect of Fatigue on Gesture Preference

A major component of physical fatigue is the part of the body that is moving. For this reason, we look at the breakdown of body part used across gestures and participants. See Table 8 for a summary.

In this study, a significant number of the gestures produced were arm gestures (78.36%). Of these arm gestures, 20.44% were coded as elbow gestures and another 21.64% were coded as hand gestures. These account for over half of the arm gestures (42.08%). In this study, gestures that were coded as elbow gestures could've included gestures where the point of rotation was the elbow (therefore the forearm was moving without the upper arm), or cases like an elbow point where the actual point of rotation was the shoulder. Full body gestures in this study only accounted for 3.21%, a significant change from the nearly 23% that were produced in the first study. Like in the previous study, we see some mention of eye movements (gaze tracking and blinking) and voice interactions.

Table 8

Percentage of gesture primitives for each part of the body for Study 2

Body part used	**% of primitives using body part**
Arms (including hands, elbows, forearms, fingers)	78.36%
Legs (including feet, toes, knees)	12.22%
Full Body	3.21%
Head	3.01%
Eyes (gaze tracking, blinking)	1.80%
Voice	1.40%

To more closely examine how fatigue affects user preference, we compared how often participants' rankings for their preferred gestures changed after repeating the gestures for 10 seconds each. On average the participants performed each gesture about 8 times (*SD* = 4.34) during the 10 second interval. The results are below in Table 9. Before repetition the position of the most preferred gesture was on average at position 2.18 (*SD* = 1.41), whereas after repeating the gesture for 10 seconds, the average most preferred gesture was at position 2.46 (*SD* = 1.43). Therefore, repeating the gesture and inducing fatigue changed user preference towards gestures that may be seen as less discoverable (i.e. elicited later in the process).

In this table, it is clear that participants often changed the rankings of all of the gestures they had defined after they were asked to repeat their gesture continuously for 10 seconds. The gestures that saw the least amount of change were those that belonged to *page* and *scroll*, which also had some of the highest max-consensus scores. In contrast, *return to main menu*, which had the lowest max-consensus of only 40 %, also had the highest change in preference across surveys.

When participants were asked the reasons for their rankings, 70% of them said that they ranked the gestures based on their level of comfort in performing the gestures. This response did not vary between the first and second survey, meaning that comfort was important and considered by participants even before they were asked to repeat all of the gestures for an extended period of time.

4.3.3 Relationship Between Discoverability and Fatigue

Table 9 also contains the results for the number of preferred gestures that were performed first, both before and after repeating the gesture for 10 seconds. Looking more closely at the most preferred gesture in this table, it seems that participants were very likely to change their gesture preferences away from the first gesture they performed for each referent. There are some exceptions. A few participants decided that the first gesture they defined was actually their most preferred after performing the gestures repeatedly. This happened in the case of the Page referent, where one participant moved away from their first gesture performed, while another one, after repeating the gesture, determined that he/she preferred the first gesture they performed more than the original first choice.

Table 9

Changes in the most preferred gesture across surveys

		First Survey			Second Survey		
Task	Percentage of Preferred Gesture Changes from First to Second Survey	Number of Preferred Gestures Performed First	Total Preferred Gestures	Percent	Number of Preferred Gestures Performed First	Total Preferred Gestures	Percent
Activate screen	46.67%	9	15	60.00%	5	15	33.33%
Page*	18.75%	8	16	50.00%	8	16	50.00%
Return to main menu	67.00%	6	15	40.00%	4	15	26.67%
Scroll	20.00%	10	15	66.67%	7	15	46.67%
Select	46.67%	7	15	46.67%	6	15	40.00%
Undo	40.00%	4	15	26.67%	5	15	33.33%

Note. Participants often changed their most preferred gesture after repeating every gesture for 10 seconds. This was still true for gestures that were performed first for each referent.

* One participant ranked two gestures as their most preferred for Paging.

To assess whether this phenomenon was statistically significant, we conducted an n-way ANOVA with unbalanced groups to identify whether the gesture number, task or participant had an effect on the changes in ranking for all 438 gestures made across all participants. The changes in ranking provide an indirect measure of fatigue, while the order in which the gesture was elicited is an indication of discoverability, allowing us to explore whether there is a relationship between discoverability and fatigue. The gesture number represents the order in which the gesture was elicited by the participants each time. For example, if the participant elicited a "swipe", "tap", "step" and "kick" for page right, swipe would be gesture number 1, tap gesture number 2, etc. We found no effect for task (F (14, 412) =.18, p = 0.969) or for participant number (F (14, 412) = 0.24, p = 0.998). However, we found a statistically significant (F (14, 412) = 0.24, p < .01) effect for gesture number and change in rankings. There is an inverse relationship between discoverability and fatigue, with gestures that are ranked high initially dropping in ranking after repetition and gestures originally ranked lower increasing in ranking.

4.4 Discussion

4.4.1 Discoverability and Conflicting Gesture Sets

As can be seen in Table 6, there are instances in which elicitation studies produce ambiguous and conflicting gesture sets if one were to choose the most agreed upon gesture for each referent. This is because each referent is treated independently in the study design. One possible approach to resolving this issue, is to look at gestures that have the second-highest consensus scores for a referent, and in cases where there is little difference between the gesture with the highest and second-highest consensus to pick the gesture with the second-highest consensus score. In this study, this approach is possible and produces a non-conflicting gesture set, but it may not work for all studies and contexts (e.g. if there are more referents, or if the referents presented are very similar or draw on the same cultural metaphors).

Another approach is to use a framed guessability approach to elicitation studies (Cafaro et al., 2014, 2018). This approach extends the concept of embodied allegories (Cafaro, 2012) to constrain the metaphorical reasoning space for possible gestures by introducing a frame (i.e. a "scenario"), thereby reducing the number of valid options (Cafaro et al., 2014). In their work, Cafaro et al. found that users preferred gesture sets constrained by embodied allegories rather than gesture sets created through traditional gesture elicitation methodologies (Cafaro et al., 2014). Cafaro et al. then extended this approach with priming and found that gesture sets created with framed guessability were likely to all share a common frame as set up by the priming, and therefore more discoverable even when there was no reference to the frame ("the scenario") in situ. They are also less likely to be affected by legacy bias than those generated via traditional elicitation methods. However, the authors suggest that the frame chosen during the elicitation phase have clear connections to the in situ scenario, meaning that elicitation studies conducted in this way still do not provide is with generalizable design principles.

Like Cafaro et al., we also provided participants with a "scenario", however our scenario was the one we were designing for, not a scenario meant to elicit creative output like the "funhouse" example. It is also not clear whether gestures produced with this approach are fatiguing or not based on the existing studies. One could imagine a scenario such as the "gym" to elicit much more fatiguing gestures by the nature of the activity that users engage in when at the gym. More research would need to be done using framed guessability to understand the connection between the scenarios and the effects they have on the gesture sets produced.

4.4.2 Measuring Fatigue

In this study our initial aim was to quantitively measure fatigue and compare that to the qualitative fatigue measures derived from our survey to see if they agree, and if not, which gestures might contain differences between measured physical fatigue and perceived fatigue. Due to the fact that we had to discard the GSR data because of individual differences and a lack of baselining, this was not possible. Other researchers have started to look at quantitative measures of fatigue (J. D. Hincapié-Ramos, Guo, & Irani, 2014b; J. D. Hincapié-Ramos, Guo, Moghadasian, et al., 2014; Jang et al., 2017). The measure developed by Jang et al. specifically aims to take perceived and cumulative fatigue into account (Jang et al., 2017). But all of these measures are based off of shoulder joint rotation, as this is the joint in the arm that's most likely to become fatigued first (Law & Avin, 2010).

In this study, we saw a significant increase in the use of arm gestures over full body ones (78% compared to 61% in the first study). The increase in arm gestures could have be due to the fact that this study used a small 21" display standing on a table instead of the 63" wall-sized display, as this would be consistent with observations from the first study in which participants used larger gestures for more or larger objects and smaller gestures for smaller objects. From Lenman et al., we know that gestures that require hand or arm movements without support are likely to be difficult for users to repeat or perform for extended periods of time without fatigue (Lenman et al., 2002), so being able to accurately measure fatigue becomes critical. This also highlights the need to understand whether basing measures of fatigue off of shoulder joint rotation is appropriate, since more than half of the gestures in this study were hand or elbow/forearm gestures. That means the shoulder may have had no rotation component, rendering all of the existing quantitative measures inaccurate without modifications. In the next study, one of the features to more closely analyze is the primary joint of rotation (the joint most likely to fatigue) for each gestures to answer this question.

4.4.3 The Tension Between Discoverability and Fatigue

Much of the research focuses on either measuring the fatigue of gestures (J. D. Hincapié-Ramos, Guo, & Irani, 2014b; J. D. Hincapié-Ramos, Guo, Moghadasian, et al., 2014; Hoggan, Nacenta, et al., 2013; Hoggan, Williamson, et al., 2013; Jang et al., 2017; Kölsch et al., 2003; Ruiz & Vogel, 2015) or on finding easily discoverable or guessable gestures (Cafaro et al., 2014, 2018; Grijincu et al., 2014; Vatavu, 2012; Wobbrock et al., 2005, 2009) for touch or full-body interactions, however, none of the research so far has looked at the relationship between the two. In previous studies, we found that users do not often prefer what we would think of as the most discoverable gestures – those that they are likely to perform first. In this study, we take this finding one step further by exploring why the most discoverable gestures are not preferred, and we find that comfort is one of the most important factors to users in defining gestures, that users do not properly assess how tiring a gesture is prior to repetition, and that the most discoverable gestures often become even less preferred after repetition, showing an inverse correlation between discoverability and fatigue.

This creates a tension between designing a gesture set that is easy to discover, especially in public spaces, where extensive training is not an option, and designing gestures that might be less discoverable, but allow for prolonged interaction. While the consumed endurance model estimates that users can only hold their arms up for 90 seconds before experiencing fatigue (Bachynskyi, Palmas, Oulasvirta, Steimle, & Weinkauf, 2015; J. D.

Hincapié-Ramos, Guo, Moghadasian, et al., 2014), this study shows an impact of fatigue on gesture preference after just 10 seconds of repeated motion. This makes designing gestures for public displays challenging.

One possible solution in full-body interaction is to move users away from simply using arm gestures, to using a combination of arm, speech and full-body gestures, providing more rest time in between arm movements. However, then the challenge lies in communicating to users, in a few moments when they walk up to an interactive display, not just that the display is interactive, but which gestures map to which actions in the particular application. We revisit this point in Chapter 6, after looking more closely at some other factors that might influence user preference in the next study.

4.5 Summary

In this study we explored discoverability and fatigue of gestures and the relationship between the two. Our results indicate that:

1) **Discoverability: Is the first gesture defined the one that users prefer on their initial ranking?** On average the most preferred gesture is still the 2nd or 3rd, even on initial ranking, which is consistent with the previous study. This means that the most discoverable gestures are not the most preferred.
2) **Fatigue: Do users' rankings of preferred gestures change after they have been asked to repeat the gestures for an extended period of time?** Participants often did change their rankings of their preferred gestures after being asked to repeat the gestures. This means that, even though we found that users think about fatigue from the very beginning, participants have a difficult

time assessing which gestures will be fatiguing without the opportunity to repeat the gestures.

3) **How do discoverability and fatigue relate to one another?** Discoverability and fatigue are inversely related. Gestures that are specified first drop in ranking after repetition meant to induce fatigue, while gestures that are specified later increase in ranking. This highlights a tension between designing gesture sets for discoverability and designing gesture sets that are not fatiguing.

Additionally, we modified our gesture elicitation to address one of the limitations in the first study by introducing a concrete scenario, with specific, non-abstract referents. However, we also used a smaller screen size that would be more appropriate for the scenario. This change may have affected the gestures produced as well, as we saw significant differences in arm and full body gestures produced between this study and the first.

The next study builds off of the study presented in this chapter by taking a closer look at how useful current measures of quantitative fatigue are. Specifically, in the next study we look at the breakdown of body parts used in gestures performed by users and the primary joint of rotation, as current quantitative measures of fatigue rely on shoulder joint rotation. In the next study, we also use a large display, to provide more insight into whether the changes in arm and full body gestures elicited were due to a concrete task or the screen size.

A significant limitation of this study was the fact that the quantitative data from the GSR sensor was discarded and therefore this data could not be compared to the qualitative assessment of the participants' perceived fatigue. A promising area for future research, that is not addressed in this dissertation, is to explore the use of current commercial wearables to measure physical fatigue and compare those measures to perceived fatigue.

CHAPTER 5

STUDY 3: UNDERSTANDING USERS' MENTAL MODELS

Building off of the previous two studies presented in this dissertation, we conducted a third study to further refine the methodology and to identify the gestural features that matter to participants most so that we may train classifiers with this added information, making them more accurate and robust to what people actually do. In previous studies, we found that asking participants to produce multiple gestures per referent increased gesture variability and that participants preferred their second or third gesture produced more than the first, which is more likely to be influenced by legacy bias. We also discovered that ease-of-use (lack of fatigue) is important to users, and that the first gesture is not only less likely to be preferred but also more likely to be identified as fatiguing by users than subsequent gestures.

However, there were limitations to previous studies. In Study 1, referent variability was low and therefore participants became bored over the course of the study, limiting their creativity for subsequent referents. In both studies, there was not enough data collected around users' mental models. In Study 1, this was due to the lack of information around iterative refinements of the gestures, gesture segmentation, and lack of clarity around what definition each participant used for favorite. Study 2, however, was focused on better understanding the relationship between discoverability and fatigue, and not on users' mental models. We also used different screen sizes and saw differences in the proportion of arm and full body gestures produced, which may have been related to the screen size, or the types of referents and the concrete scenario used in Study 2.

In this study, the primary focus is to address limitations of previous studies to better understand users' mental models (RQ 2). A secondary focus is to validate findings from previous studies and to extract design principles for walk-up-and-use interfaces (RQ 3). For RQ 2 (Which gestural features matter to users and how do they influence a user's mental model about that gesture?) the following research questions are addressed:

1) **How do we qualitatively code gesture elicitation studies to understand users' mental models?** – Over the course of the previous studies, there have been many challenges with coding gestures to better understand those features that matter to users. How do we address these challenges?
2) **What do refinements tell us about the features that are important to users?** – We saw in previous studies that users will refine their gestures throughout the course of the elicitation process. Which features change and which don't during this process?
3) **What is the primary joint of rotation for elicited gestures?** Does this change as users refine their gestures? – Study 2 indicated that users' preference is heavily influenced by perceived fatigue of a gesture, and that many gestures do not seem to have the shoulder as the primary point of rotation, rendering many quantitative measure of fatigue inaccurate without modification.

5.1 Method

For an in-depth exploration into the features that matter to users for full-bodied gestures in public walk-up-and-use contexts, we conducted a gesture elicitation study with 22 participants (12 female, 1 male, 1 gender neutral). Participants over the age of 18 (average age = 27 yrs., S.D. = 6.53) were recruited from the larger university population to participate in a one hour-long study. All participants were right handed and had various levels of comfort with technology. Half of the participants had previously used some kind of free-space gesture recognition system before (such as the Microsoft Kinect or Leap Motion) and about half (n = 11) play video games.

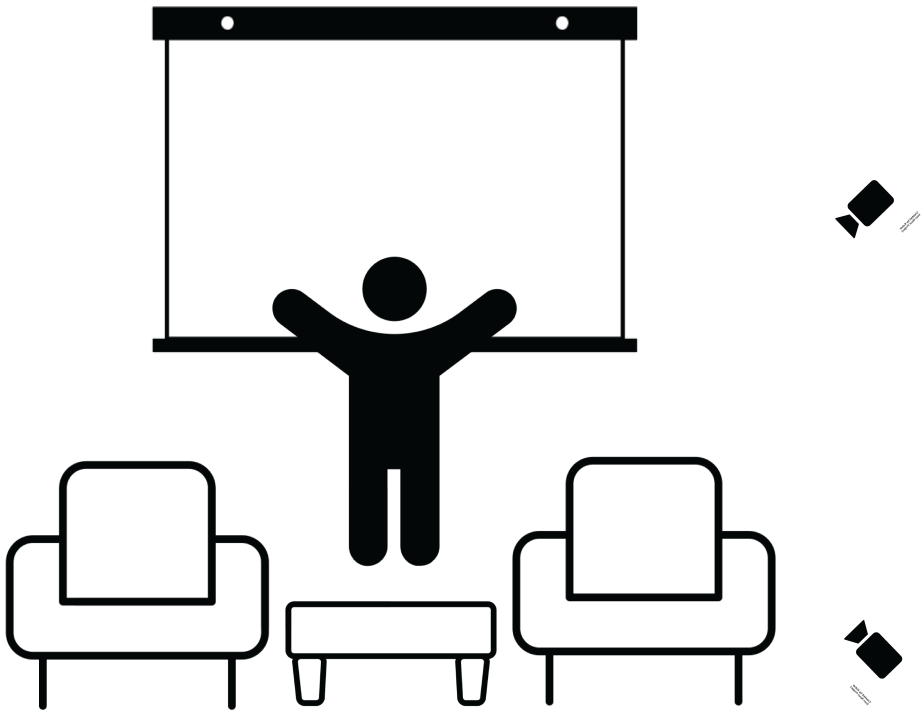


Figure 13. The study was conducted in a lab that was arranged to look like a living room. Cameras were placed at 45 degrees in front and behind the participant and they interacted with a projection display.

The study was conducted in a lab arranged to look like a living room. The study lasted approximately two hours. At the beginning of the study, participants were asked to fill out a short demographic survey (see Appendix D for the survey). The remainder of the study was split up into two parts, each of which lasted approximately 45 minutes. Participants were allowed to take a short break (5 – 10 minutes) in between. Participants were videotaped by two cameras for the duration of the study, one set up to videotape from behind the participant, and another in front of the participant. Cameras were set up on the right side of the space, one in the front corner and another in the back corner. These were positioned to be out of the way of any movement the participant may like to perform. See *Figure 13* for details of the space. All participants were compensated for their time with a 20-dollar gift card.

Unlike Study 2, this study did not focus specifically on fatigue, so no GSR sensor or other quantitative measure of fatigue was used. Because the effect of video and kinesthetic priming was inconclusive in Study 1, it was not used here. However, we did provide participants with a concrete scenario and referents, like in Study 2. The changes to the methodology related to production were maintained. For this study, a retrospective section was added to the study to gain more insights into users' mental models. The retrospective will be covered in more detail further down.

The first part of the study consisted of the gesture elicitation part, in which the participants were asked to interact with a large projection display that was approximately 6' from the chairs they were originally seated in. The surrounding area was large enough for the participant to move around in (it was approximately a 6' x 6' open area). At the beginning of the elicitation section, participants were instructed to pretend that they were in a shopping center that had a gesture-based walk-up-and-use display providing information on travel to national parks. They were told not to constrain themselves to current technological limitations and to pretend like the display could pick up anything they did. They were asked to think-aloud during the elicitation part of the study and to start by standing. During this part of the study, the experimenter had an observer packet in which they noted a short description of each gesture as it was performed, and any observations they may have made or any questions they had about the gestures so that follow-up questions could be asked during the retrospective. The experimenters also asked a few clarifying questions during the elicitation part of the study when necessarily, but this was kept to a minimum so that it didn't interrupt the participant's flow.

Each participant was shown 10 referents*,* or tasks, and asked to identify as many symbols for each referent as they could come up with. The first referent (*page left*) was used to get people familiar with the task, and therefore was not coded. The order of the remaining 9 referents was randomized and each one of the participants saw one of 4 possible referent orders. In this study, the referents were expanded to include zooming, multi-object selection and object manipulation, as well as distinguishing between scrolling speeds (slow and fast) based on results from Study 1. The full list of referents that participants were presented were: *page right, drag, scroll fast, scroll* (slow), *select, multi-object select, zoom in, zoom out*, and *deselect*.

In this study, unlike in Study 1, participants were given concrete tasks within a scenario:

> *Pretend you're in shopping center when you come across an interactive display providing information about national parks. You've always wanted to visit Yellowstone, Glacier, and Grand Teton National Parks, so you decide to learn more about what there is to see and do at these locations.*

They were also provided with actual images, in this case photos of animals and landscapes found in national parks to match the scenario, instead of abstract shapes. See *Figure 14* for an example of one of the referents participants were shown in this study. These changes were made to reduce the boredom observed in the previous study. Once the participant ran out of new gestures to perform for a particular referent, they were asked which of the gestures for that referent was their favorite and least favorite, where favorite was defined for the participant to mean the gesture which they would be most willing to perform in public. In this study, we specifically defined what favorite meant in order to account for the variability in meaning in the first study, and to ensure that participants were thinking about public displays, keeping in line with the scope of the research.

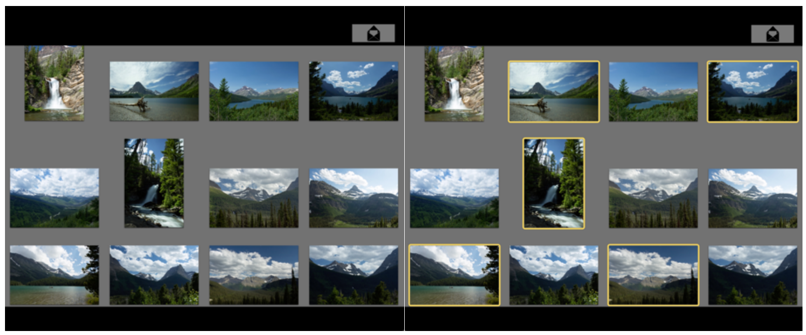


Figure 14. The "multi-object select" referent showed a set of photos arranged in a grid layout (left image) being selected one at a time until multiple images have been selected (right image). Study participants were asked to identify free-space gestures they would use to select the objects.

In the second part of the study, experimenters conducted a retrospective with the participants. During the break, the experimenter copied the recorded videos of the elicitation section onto a mac mini and loaded the videos in ELAN, making sure they were time aligned. Participants sat in one of the chairs in the space and were shown the videos of themselves performing the gestures. They were asked to rate each gesture on a five point Likert scale for discoverability, ease-of- use (to measure their feelings on fatigue), and appropriateness to the action. In some cases, the experimenter asked the participant if a gesture was a refinement of a previous one or asked any additional clarifying questions that they thought would provide necessary information. Both the ratings, and any additional information provided by the participant at this point, including whether the gesture was a refinement of a previous one, were noted in the observer packet by the experimenter.

5.2 Coding and Analysis

5.2.1 Gesture Segmentation

Pairwise coding was conducted to segment the elicitation videos by two researchers to ensure that multiple people were able to identify the start and end of a gesture sequence. Disagreements between the researchers were discussed and resolved during the segmentation process. Any movement related to the user preparing for the gesture was ignored, unless the starting point was identified to be important based on the information the user provided in the think-aloud. Many participants repeated the gestures several times during the elicitation process, so only the first time the gesture was performed was segmented. During the coding process, additional repeats were checked for inconsistencies and noted (e.g. if users switched from a 1-finger to 2-finger point). Segments of the video that provided more information or that contained instances in which the participant was exploring the gesture space were also annotated for reference.

In total, this resulted in 1117 gestures segmented (1442 total annotations, including exploratory gestures, additional information provided by the participant, and complementary gestures that didn't match the referent – for example, a gesture that was meant to *deselect* when the participant was shown a *select* referent) across 22 participants. Of the total number of participants, 4 were found to have either missing elicitation videos or the videos ended early during the elicitation process due to recording difficulties; these participants were excluded from the remainder of the analysis.

5.2.2 Gesture Coding

Of the remaining 18 participants, 10 participants were selected to be coded and their data analyzed, resulting in a total of 547 coded gestures (gestures for the first video were practice and therefore not coded). The 547 gestures resulted in a total of 660 gesture primitives. The 10 coded participants represented an even distribution of video order, participant gender, and experimenter. Three of the 18 participants are missing data from the retrospective videos, though 2 of the 3 were coded and their data was used as sufficient information was in the elicitation videos and the observer forms. Originally, we began with two researchers separately coding the gestures and comparing their coding schemes but found that achieving consensus and inter-rater reliability would be a challenge. This was due to the fact there were so many features that were being coded for, and one coder might use one feature to encode a particular piece of information, while a second coding might use a separate one. To help alleviate some of this disagreement, additional features were added during the coding process, such as the gesture path, and the starting stance. In the end, gestures were pair-wise coded by two researchers together allowing for real-time discussion of disagreements.

Features were selected based on inspiration from existing literature and informed by previous studies. As with previous studies, we began with assigning a high-level primitive to each segmented gesture (e.g. "swipe, hover, point"), side of the body used, an indication of the type or direction of motion, and similar to elicitation studies which focus specifically on hand gestures, the hand configuration. Hand configuration was coded even though evidence from Wobbrock et al. (Wobbrock et al., 2009) and Morris et al. (Morris et al., 2010) shows that the number of fingers used is unimportant in touch interactions, because for free-space interactions we may still want to distinguish between someone using a 1-finger point, pointing with a flat hand, or pointing with a fist.

Similar to previous studies, we started with a list of possible values for each feature (e.g. for gesture primitives, the list contained swipe, point, hover, kick, jump, etc.) and added to that list as gestures were encountered that could not be described by any of the current values. Since this study focuses on full body interactions, which includes gestures such as kicking, jumping, and walking around, and not just arm gestures, we coded all body parts used for each gesture. Because full body interactions often consisted of a combination of gesture primitives, we also coded the gesture primitives and their associations based on how they were performed (e.g. in sequence, simultaneously, etc.). For example, a sequence may include walking to a specific point in space and then jumping. A simultaneous gesture might include swiping with one arm up and one arm down at the same time.

In addition to the features that were coded in previous studies, we also included features to more accurately identify the path in space that each body part moved through in the users' gestures. These features were inspired by how animation software encodes animations, allowing animators to define the key parts of the movement and interpolating the rest. Likewise, the idea of key frames guided which parts of a gesture's motion to code and which parts of the motion can be trivially interpolated using the information from the gesture's sequence immediately before and after the movement. We, therefore, encoded the beginning and end points of commonly used limbs and the starting stance in cases where it was relevant.

Finally, based on previous studies, we added two additional features to the coding scheme: the primary joint of rotation and whether the gesture was a refinement of a previous gesture, and if so, which one. The primary joint of rotation allows us to determine whether existing quantitative measures of fatigue are sufficient, while the refinement allows us to more readily look at the nuances of the gesture that change over iterations, providing valuable information around the features that do and don't matter to users' mental models. For a full list of features that were coded, details on what each feature means, and examples, see Table 10.

A total of 49 gesture primitives were identified during the coding process. Once all of the gestures were pairwise coded, the experimenters went through all the coded gestures together, and resolved any remaining discrepancies. Some of the gesture codes were further refined. For example, similar gestures that were coded as discrete primitive sequences (like touch and tap) were collapsed into one higher-level gesture primitive (e.g. tap). In another example refinement, a sequence of gestures was collapsed into a single gesture primitive with additional information captured into a different feature (such as collapsing all drawing gestures into one gesture primitive with the path details containing the thing that was drawn, e.g. an "X"). Generally speaking, such refactoring was focused on making the gesture coding more consistent and where possible, smaller without loss of detail. 38 gesture primitives remained after consolidation. For the resultant list of gesture primitives and their descriptions, see Appendix E.

Table 10

Features used to qualitatively code the data for each participant's gestures.

Feature	Definition	Examples

Gesture Primitive	High level label of gestures. There were a total of 37 gesture primitives identified.	Point, swipe, kick
Gesture # in Video	When was this gesture specified for the video?	1, 2, 3, 7
Gesture Type	Some gestures are made up of a sequence of gestures, or multiple movements may happen simultaneously or repeat. If it was a sequence, the start and end was noted.	Repeating, Sequence, Simultaneous
Refinement?	Is this a refinement of a previous gesture and if so, which one?	1, 3, 2 (reference to the gesture being refined)
Body Part Used	Which part of the body was used to perform the gesture?	Arm, hand, leg, head, full body
Gesture Direction	What was the primary direction of travel?	To an object, forward, back, right, left
Which Object?	If the direction of travel was towards an object, which object?	To empty space, to selected object

Gesture Path	Was it a straight path or did it arc in any way?	Straight, arc, circular
Starting Stance	Was there a non-neutral starting stance that mattered?	Shoulder width, right foot forward
Palm Position Start / End	Direction of the palm relative to the body.	Forward, back, in (towards body)
Hand Config. Start / End	What configuration is the hand in?	Fist, flat hand, 1-finger
Forearm Start / End	Position of forearm relative to elbow.	Up, forward, partially down
Upper Arm Start / End	Position of upper arm relative to shoulder.	Up, right, left, partially up
Point of Rotation	The joint that is most fatiguing (i.e. shoulder movements are more fatiguing than elbow movements, which are more fatiguing than wrist movements) that moves for the gesture.	Shoulder, elbow, knee
Relative / Absolute?	Is the gesture relative to the body or absolute in space?	Relative, absolute
Favorite / Least favorite	Was this the most or least favorite for the referent?	favorite, least favorite n/a

5.3 Results

In this section, we present the results of the qualitative coding. Because one of the goals of this study was to address limitations of previous studies and validate results, in addition to conducting a more detailed analysis, the results section is broken down by overall research questions. The questions presented at the beginning of this chapter, specific to this study, are discussed in the context of the larger research questions.

5.3.1 Modifications for Legacy Bias

In this section we give an overview of the effect of production on the gesture elicitation process. Consistent with previous studies, participants produced on average 5 gestures per referent (SD = 1.74). Also confirmed in previous studies, the most discoverable gesture is not the most preferred. The median position of the favorite was 3 (SD = 2.07) and the median position of the least favorite was 4 (SD = 1.68). Some participants had multiple favorites or least favorites per referent. Unlike the rest of the participants, P7 stood out because they generally preferred the first gesture they specified per referent (for all but 2 of the referents).

Table 11 shows the means and standard deviations for the favorite, least favorite, and no preference gestures (those which weren't rated as either favorite or least favorite). Since participants rated the gestures, not the gesture primitives, we removed duplicated gesture ratings across primitives, leaving 547 gestures to analyze. A one-way ANOVA with unbalanced groups shows that the three groups are statistically significant for all three user ratings (guessability ($F(2, 454) = 30.01$, $p < .001$), ease-of-use ($F(2, 454) = 33.32$, $p < .001$), and appropriateness ($F(2, 454) = 37.56$, $p < .001$)).

Table 11

Means and standard deviations for the guessability, ease-of-use and appropriateness ratings across user preference

User Preference	Mean Guess.	St. Dev. Guess.	Mean Ease	St. Dev. Ease	Mean Approp.	St. Dev. Approp.
Favorite (n =97)	4.08	1.17	4.63	0.79	4.54	0.77
Least Fav. (n=89)	2.62	1.37	3.33	1.34	3.04	1.34
No Pref. (n = 361)	3.58	1.28	4.13	1.07	3.83	1.21

Table 12

Percentage of gesture primitives for each part of the body for Study 3

Body part used	% of primitives using body part
Arms (including hands, elbows, fingers, etc.)	68.33%
Legs (including feet, toes, knees)	9.70%
Voice	8.33%
Gaze	5.30%
Full body	4.55%
Head	3.33%
Brain Machine Interfaces (BMI)	.45%

Note: Only one person mentioned ASL across referents, two participants mentioned blinking, and three participants mentioned BMIs.

As with previous studies, we take a look at the breakdown of the types of gesture performed to understand whether the changes in production increased the variety of gestures produced. In this study, the variety of gestures produced is much more consistent with findings from the first study. Arm and hand gestures were still the most common, at about 68%, however full-body gestures (e.g. walking, leaning) only made up 4.6% of the gestures produced in this study. Gestures using the legs and feet were similar, at 9.7% this time, compared to about 7.5% in the first study. In this study, unlike previous ones, there were a couple of participants that mentioned brain-machine interfaces.

Unsurprisingly, head gestures were not used at all for *selection* (either single or *multi-object selection*), but they were used for *deselect*. Leg gestures weren't ever mentioned for the *drag* referent, but were for all others. Voice was mentioned for every single referent by at least one participant. Table 12 shows the breakdown of the body parts used across all gestures, all participants and all referents.

We also look at the agreement across participants for this study. Overall, all the referents had very high max-consensus scores and many referents had high consensus-distinct scores as well (Table 13 shows the max-consensus and the consensus-distinct scores for each referent), indicating that legacy bias still plays a significant role and leads participants to specify the same few gestures for each of these actions. This is confirmed by participants (P3, P13, P15) explicitly identified that they were drawing from interactions with smart phones.

Table 13

Max-consensus and consensus-distinct scores for each referent

Referent	Gesture with Highest Consensus	Max-Consensus	Consensus-distinct
Page Right	Swipe	100%	0.421
Drag	Move (arm/hand)*	90%	0.882
Scroll Fast	Swipe	90%	0.706
Select	Point	90%	0.667
Multi Object Select	Point	90%	0.579
Zoom In	Expand	80%	0.684
Zoom Out	Pinch	80%	0.500
Scroll	Swipe/Slide	60%	0.600
Deselect	Swipe/Tap	60%	0.400

* Move in this case refers to moving one's hand over to the email icon shown on screen to drag the item over. Sometimes, this motion was preceded by a "grab" type movement, but not always.

Similar to Study 2, looking at just the gestures with the highest consensus does not necessarily provide a unique non-conflicting gesture set for all referents, however, there is much less overlap in this case than in Study 2. Only one participant, P18, mentioned that they cared about how confusable the gesture was for this referent compared to the ones for other referents but most participants did not take this issue into account when defining gestures. Note that this study does contain referents that are similar and may be expected to map to similar gestures as well. Table 14 shows the top 3 gesture primitives that are most common based on the max-consensus metric for each referent.

Table 14

The top 3 most common gesture primitives for each referent in Study 3 and their max-consensus scores

Referent	Gesture	Max-Consensus
Zoom In	Expand	100%
	Voice	50%
	Tap/Point	40%
Page Right	Swipe	100%
	Step	50%
	Voice	50%
Drag	Move (arm /hand)	90%
	Point	60%
	Grab	60%
Scroll Fast	Swipe	90%
	Nod	50%
	Gaze	40%
Select	Point	90%
	Tap	70%
	Voice	60%
Multi-Object Select	Point	90%
	Voice	70%
	Tap	60%

Zoom Out	Pinch	80%
	Voice	50%
	Rotate (Body or Hand)	40%
Scroll	Swipe	60%
	Slide	60%
	Voice / Nod	50%
Deselect	Swipe	60%
	Tap	60%
	Voice	50%

5.3.2 A Closer Look at the Features That Matter

To better understand the features that matter to participants when specifying a gesture, we take a closer look at the refined gestures and the changes between the original gesture elicited and the refinement. We also take a closer look at the primary joint of rotation. As we saw in previous studies, users tended to reduce the magnitude of their gestures over time and current quantitative measures of fatigue are primarily developed with the shoulder joint as the primary point of rotation.

Out of the 547 gestures coded across our 10 participants, 57 of them were refinements (10.42%). We began by evaluating whether the ratings change between the refinements and the original gestures being refined. Table 15 shows the average usability, ease-of-use and appropriateness ratings for both the refinements and the original gestures. A one-way ANOVA showed no significant difference for appropriateness (F (1, 98) = 0.72, p

= 0.398) and ease-of-use (F (1, 98) = 1.72, p = 0.193), but was borderline non-significant for guessability (F (1, 98) = 3.88, p = .052).

Table 15

Means and standard deviations for the guessability, ease-of-use and appropriateness ratings for refinement gestures, the gestures being refined and for the entire set of elicited gestures as a comparison.

	Mean Guess.	St. Dev. Guess.	Mean Ease	St. Dev. Ease	Mean Approp.	St. Dev. Approp.
Refinement	3.63	1.24	4.16	1.01	4.12	1.15
Original	3.98	1.14	4.31	0.91	4.25	0.96
All	3.50	1.36	4.08	1.15	3.90	1.26

Additionally, of the combined refinements and originals, 29 were listed as favorite gestures (25.67%), whereas only 16.90% of all gestures are listed as favorites. Overwhelmingly, participants seemed to refine gestures for which they already had a preference. A two-sample two-tailed t-test indicates this difference is statistically significant ($p < 0.01$).

Another aspect explored was the primary joint of rotation, as we know that fatigue measures are currently based off of the fatigue experienced by the shoulder joint and that fatigue matters to users. Even in this study, some participants mentioned that the level of fatigue a gesture causes is an important consideration during the elicitation process and for this reason participants wanted to specify gestures that were “more intimate” (P3) or that “take very little effort” and are inconspicuous (P7).

To explorer this area further, we looked at how often each of the joints was actually used. As can be seen in *Figure 15*, current measures of fatigue can only provide a quantitative assessment for less than half of the gestures specified by participants. In the figure, gestures that are marked with N/A for the joint used are gestures that do not have movement associated with the gesture primitive (e.g. dwell).

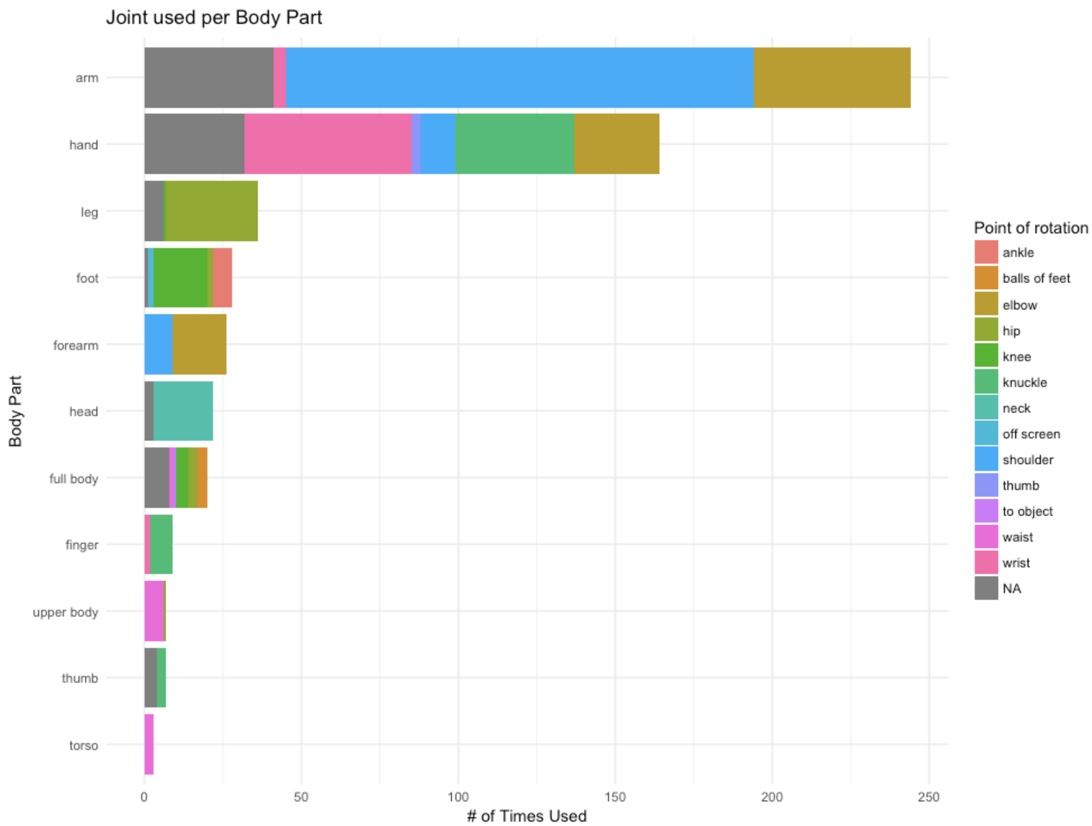


Figure 15. Overview of the primary joint of rotation per body part across participants. An N/A means that there was no movement of any joint for the gesture (e.g. point and dwell gestures don't usually have joint movements associated with them)

Finally, we analyzed a subset of the features that changed between the refinement and original gesture. The frequency of each feature changing can be found in Table 16. For the scope of this work, we did not closely look at the detailed paths that the arms took, so we did not analyze changes in the start and end position of the forearm and upper arm.

Table 16

Frequency of each feature changing across refinements

Feature	n (%)
Palm Direction	42 (73.68%)
Hand Configuration	34 (59.65%)
Gesture Direction	18 (31.58%)
Point of Rotation	16(28.07)%
No. of Gesture Primitives	14 (24.56%)
Body Part	13 (22.81%)
Gesture Path	12 (21.05%)
Side of the Body	6 (10.53%)

All refinements were on arm gestures, not on gestures that were full-body or leg gestures and nearly all of the refinements happened immediately after the gesture they were refining. The most likely features to change between the refinement and original gesture were the palm direction and hand configuration. Gesture direction also varied but did not have any discernable patterns in the change. Gesture path, however, often changed from a straight path to an arching or circular path. The number of gesture primitives changed as well. In some cases, users added extra primitives, and other times simplified. Adding primitives happened in cases where the gesture's intent became more specific. For the side of the body, in all but 1 of the 6 cases, the change was between using both sides of the body or only one side.

Out of the 30 cases in which either body part or point of rotation changed, 10 of those cases are instances in which both changed. In all but 5 of the 30 cases, the movement became smaller (e.g. using the hand instead of the entire arm, or changing the point of rotation from the shoulder to the elbow or even the wrist). In 2 cases, the user changed the gesture to use a different finger (e.g. thumb instead of pointer finger), and in 3 cases, the movement became larger.

5.3.3 Users' Mental Models and Design Principles

In this study, additional evidence emerged for the themes presented in Study 1.

Theme 1: Similarities Across Selecting, Scrolling, and Paging

In this study, the palm direction was often maintained constant across participants though there were a few exceptions (e.g. P2). In study 1, we saw several indications of users swiping both left and right without changing the direction of the palm.

Theme 2: Gestures on Body, Gestures in Space

Similar to the first study conducted, some participants (P2, P9, P14) mapped the screen in front of them to a horizontal plane, often specifying gestures as if they were interacting with a touch table instead of a display on the wall in front of them. This continued even when the researchers reminded the participant that the display was in front of them. Other participants wanted to walk up to the screen and touch or tap on the screen directly (P3, P7, P15), even though they could not reach the entire screen due to its size. For leg or full body interactions, many participant mapped the screen to the floor (P2, P14, P15, P17, P19).

Theme 3: Gestures for Size and Speed

As can be seen in Table 13 and Table 14, *scroll fast* had a 90% max-consensus score for swipe, while *scroll* (which is a slower movement) had slide and swipe tied at 60% (slide and swipe only differ on the speed of movement). Every participant produced either one or the other for this referent (2 participants mentioned both). This is consistent with findings from Study 1 that participants prefer faster arm motions to map to faster movements on screen as well (P3, P9). For example, P9 wanted the speed of the movement to influence how quickly the movement happened on screen, specifically called out gestures that they did not like because they thought the motion was too small (e.g. P9 mentioned that 1-finger is good for actions on a single object, but not many).

One participant, P3, mapped speed to the angle of rotation of their hand. Another participant made the same gesture with a different joint as the primary point of rotation to specify speed (P19).

In this study, the size of the images users interacted with was kept constant and in all cases where multiple images were presented, they were shown in a grid layout. This meant that we could not observe any variability in the magnitude of the gestures for the number of objects or size of objects present. However, P7, explicitly mentioned that the size of the display influenced the magnitude of the gesture.

Theme 4: Gestures for Parallelizing Data Exploration

In this study, we did not have any tasks that required users to parallelize data exploration, and all participants were right handed. Unlike in Study 1, where whether the user used the right or left side of their body was influenced by the location of the object they were interacting with, in this study, nearly all of the interactions were performed with the right side of the body, with 414 gestures performed on the right side, 100 performed with both sides, and 48 performed with the left side of the body. Note, however that the selection in the *select* referent, *page right* and *drag* were all on the right side of the screen. *Zoom in*, *zoom out, scroll, scroll fast* were either centered on screen or had no interaction that forced participants to interact on any particular side of the screen. *Deselect* and *multi-object select* had the same objects selected and were split across both sides of the screen with more objects on the left that the participant was interacting with.

Theme 6: Concerns about Ambiguity

While participants still had concerns around ambiguity in this study similar to previous ones, especially around gestures using the head, or gaze, another source of confusion and ambiguity also became apparent. Participants were often likely to specify complementary or identical gestures for referents that had complementary actions (e.g. select and deselect) or for which the action was the same, but the direction of motion was different (e.g. scroll up /down, page right/left, zoom in/out). For example, many participants wanted to select and deselect an individual image by pointing at it. However, many participants (e.g. P3, P14) had difficulty remembering which complementary gesture should map to which action (e.g. does swiping from left to right mean that the participant would see the image to the right of the current one on screen or the one to the left? Similarly, does a pinch gesture result in the image zooming in or out?). This resulted in participants specifying both complimentary gestures for the same referent.

5.4 Discussion

5.4.1 Methodological Changes and Coding Full-Body Gestures

To improve the quality of the data collected in this study and to make coding the gestures less ambiguous we made several methodological changes, such as using concrete referents and introducing a scenario, introducing a retrospective, and changing the way the qualitative coding was conducted. As in Study 2, in the study presented in this chapter, we used concrete primitives engrained in a scenario. As a result, the boredom effect that was present in study 1 was not observed here. This is likely due to the larger variability of the referents and the fact that participants had concrete objects to interact with. As in study 2, our concrete scenario differs from the type of scenarios that Cafaro et. al use in their framed guessability methodology (Cafaro et al., 2014, 2018).

The addition of a retrospective allowed us to ask for the ratings of ease of use, appropriateness and guessability without breaking creative flow (Csikszentmihalyi, 1997) and it allowed us to more accurately identify refinements and gauge user preference. Retrospectives also allowed us to confirm certain ambiguities around which gestures users thought were unique, not just for refinements, but also to filter out exploratory gestures or gestures that were not meant to be considered from the participant's perspective. This significantly reduced the difficulty researchers had in identifying the start and end of gestures to segment them. Even so, participants started out with very defined starts and ends to their gestures, and as they repeated a given gesture the gesture would become more and more sloppy, blending the start and end of the gesture together as it was repeated.

In this study, we used a similar sized screen to the one used in Study 1, and the gestures broken down by body part used was also more similar. We will take a closer look at these similarities and differences across the three studies in the next chapter.

Over the course of the three studies we managed to reduce the number of unique gesture primitives from 133 in the first study to 38 in study three (Study two had 72). This allows us to more appropriately group similar gestures together and get a better sense of max-consensus and consensus-distinct scores across participants. Unlike most studies, we also split gestures up into primitives when the gestures used to accomplish a task were complex (for example, the user moving to a position in space then swiping to select an item would be considered two gesture primitives – a move and a swipe). In many studies, these are considered as one gesture and coded as such. This increases the variety of gestures, but prohibits a thorough analysis of what users consider similar movements. For example, swiping to scroll, and swiping away an item to deselect it, after moving to the position in space the object is at, are similar gestures to a user, even though the latter is part of a compound gesture meant to perform one action. Making the change to break down an entire set of movements into gesture primitives does increase the complexity of the analysis however, and makes calculating a max-consensus score much more difficult as preferences may exist for different gesture primitives at different stages of the movement.

We were also able to collapse nuances that did not matter to users in the gesture primitives, and identify more detailed features that could matter, such as the primary joint of rotation. We also saw less variety in the later studies, which could be due to increased prevalence of touch and gesture interfaces (higher legacy bias), more concrete tasks, differences in participant demographics (the last two studies were conducted on university campuses), a lack of priming or other unidentified factors.

Coding qualitative gestures with this many degrees of freedom is still challenging however. Attempts at interrater reliability were time consuming and resulted in a significant amount of discussion, specifically for features that were more detailed. Ultimately, it was more efficient to pair-wise code the entire data set than to split up the participant data and code it separately. For this reason, and because we believed we were able to achieve near saturation (Glaser & Strauss, 1967) at 10 participants, we stopped before coding the full set of participants that participated in the study. We look more closely at sample size and saturation in Chapter 6.

5.4.2 What Refinements Tell Us About User Preference

In our analysis, we noticed that users often refined gestures that they already preferred, with gestures that were refined containing a much higher percentage of favorites than the full set of gestures elicited from participants. Refinements usually happened immediately, so as soon as participants found a gesture they thought they might like, they put the effort into exploring refinements. All refinements were focused on arm or hand gestures, not on gestures that were full-body or leg gestures. One possibility for this is that arm and hand gestures are more likely to suffer from legacy bias, as touch and desktop interactions use exclusively hand movements and many Kinect interactions also used arm gestures for navigation and selection tasks. Another reason for refinements focused exclusively on arm gestures is that arms inherently have more degrees of freedom than other parts of the body, allowing for more possibilities to be explored.

Our results also indicated that some of the features that are used to train classifiers now: palm direction and hand configuration especially, seem to contain a large amount of variability and are not distinguishing characteristics for users. The finding for hand configuration is consistent with findings that users do not place importance on the numbers of fingers used in touch interaction (Morris et al., 2010; Wobbrock et al., 2009). Hand configuration is only meaningful when the hand configuration itself is the gesture (for example, a thumbs up gesture, or an L-shape). These, may therefore, not the best features to use as discerning characteristics when designing gestures and training classifiers. Recognizers also need to be trained with data that allows for some variability in gesture path, as straight and arching paths are used interchangeably by users. For example, one participant may swipe from the top left to the bottom right in an arc motion the first time, and swipe from left to right across the body using a straight path a second time. In another case, a user may draw an "X" from the top right the first time around and from the top left the second time around.

In comparing the user ratings for guessability, ease-of-use and appropriateness of refinements with the original gestures being refined, we found there was no statistically significant difference. The ratings for guessability were borderline non-significant, which may indicate there is a small effect there that cannot be detected with such a small sample size. However, refinements almost exclusively contained changes in the magnitude of the gesture between the original and the refinement, getting smaller through changes to the part of the arm that was used, or by changes to the primary joint of rotation. The changes in the magnitude of the gestures would indicate, then, that participants did take fatigue into account during the refinement process, but that it either was not the primary motivating factor or they were simply unaware that fatigue was driving this decision. The work of Liu and Thomas, who found that the more fatiguing a gesture is perceived to be, the less appealing it is to participants (X. Liu & Thomas, 2017), supports the theory that users were just not aware that fatigue was driving their decision.

It is possible that other ratings of perceived exertion, such as the NASA-TLX (Hart & Staveland, 1988) or the BORG CR10 (Borg, 1998) would uncover perceived exertion differences that our Likert style questionnaire did not. The BORG CR10 has already been used to assess perceived fatigue in arms and hands and has been shown to correlate well with objective measures of fatigue (Jang et al., 2017). One avenue for future research is to evaluate whether fatigue is the largest motivator for refinements, or if there are other factors that have not been identified or studied yet.

5.4.3 Limitations of Current Quantitative Measures of Fatigue

Even though results indicate that half of arm gestures have the shoulder as the primary joint of rotation, we also find that users modify their gestures during refinements to use other, less fatiguing, joints and smaller movements. From the previous study, we know that fatigue and comfort play an important part in user preference, so it becomes imperative that there are measures of fatigue that are more robust and can be used for the wide variety of gestures produced, especially those that are preferred. When taking into account that only 68% of gestures for this study were arm gestures to begin with, and then half of those had the primary point of rotation as the wrist or elbow, that leaves only about 34% of gestures that can be accurately assessed with current measures, such as those by Hincapié-Ramos et al. (J. D. Hincapié-Ramos, Guo, & Irani, 2014b; J. D. Hincapié-Ramos, Guo, Moghadasian, et al., 2014), or by Jang et al. (Jang et al., 2017). Refining quantitative measures of fatigue to allow for evaluating a broader range of gestures is a rich area for future research.

5.5 Summary

Looking back at our research questions for this study, we can summarize our findings as follows:

1) **How do we qualitatively code gesture elicitation studies to understand users' mental models?** In this study, we were inspired by various means of coding and discussing gestures, and refined our coding scheme from previous studies, to code a set of 16 features, including refinements that providing a rich amount of information about users' mental models. We were able to more accurately segment gesture data, and gather information about user preference.
2) **What do refinements tell us about the features that are important to users?** Refinements highlighted that gesture path, gesture direction, hand configuration and palm direction experience a lot of variability within and across users and therefore are not as meaningful to users, highlighting a mismatch between users' mental models and current gesture recognizers. Refinements also show us that users often minimize their gesture to make them less fatiguing, pointing towards designing systems with gestures that utilizes smaller parts of the arm instead of full-arm movements.
3) **What is the primary joint of rotation for elicited gestures?** While the primary joint of rotation is still the shoulder for many gestures, this is likely caused by legacy bias. Especially, since we see in users' refinements that the primary joint of rotation often changes to the elbow or wrist. This highlights a need for more robust measures

of fatigue and for gesture recognition systems that can recognize smaller movements, even with large public displays.

A few interesting areas for future research were highlighted in this study. One is to more closely analyze how well user preference lines up with agreement. We see that, most often, gestures that have high agreement are strongly affected by legacy bias. We also see that user do not necessarily prefer those gestures that are elicited first, and that are most likely to have high discoverability and agreement. Finally, we see that favorite gestures, which are on average the third gesture that people produce, have much higher guessability, ease-of-use (low fatigue) and appropriateness ratings than other gestures. This points to the possibility that there might be a mismatch between gestures that have high consensus, and those that are preferred by users.

We also did not closely analyze the forearm and upper arm keyframes that were qualitatively coded, and which might provide additional insight into the variability of user gestures that need to be collected for accurate and robust training of gesture recognition classifiers. Finally, a logical next step would be to create a gesture sett based on the results of this study, develop a system with a recognition engine based off this gesture set and evaluate how well it does in a walk-up-and-use scenario.

CHAPTER 6

DISCUSSION

In this chapter we discuss the differences and themes that emerged across the three studies, highlight limitations of the research presented and highlight future work.

6.1 Differences Across Studies and Associated Effects

There were several potentially confounding factors across the study that makes comparison across the studies difficult. In this section, we review what those factors are and highlight differences across the studies. Study 1 was conducted in mid-2012, Study 2 was conducted in early 2015, whereas Study 3 was conducted in early 2014, but analyzed after Study 2. Not only were there differences across the three studies, but during the course of this timeframe, the ubiquity of touch, gesture, and voice interaction technology changed, which influences users' legacy biases.

First off, the demographics between study 1 and study 2 and 3 differed. Study 1 was conducted with local participants from the greater Seattle area, encompassing a wider variety of education levels and age groups. Studies 2 and 3 were conducted with a university population, which will have a much smaller variability in age groups and education levels, but a higher cultural variety.

We also made several modifications between studies, specifically to address limitations of prior studies. In study 1, we leveraged kinesthetic and video priming, but because we were not able to detect a strong effect due to priming we removed it for the previous studies, and focused on production, which remained constant throughout the three studies. Removing priming from the study methodology may have had an impact on the variability of gestures produced.

We also changed the concreteness of the tasks and introduced a scenario after Study 1 to decrease the boredom effect we saw. This change did have the effect of reducing boredom, but could have potentially reduced the variability of gestures produced as well. Finally, the tasks were different across the three studies, with some similarities (all studies included *scrolling, paging and selection* referents). Study 1 included text, not just object manipulation and was geared towards facetted browsing, most visible through the 3-column layout referents and the large variability in layouts that were being tested for browsing and navigation tasks. Study 2 contained a very small set of referents, some of which were focused on less common tasks (*undo, return to main menu*). Both Study 2 and 3 had a smaller variability in layouts, as each referent was meant to represent a different part of interacting with the same application within the same scenario. These changes were made to be more in line with the scenarios introduced.

Similarly, unlike study 1 and 3, Study 2 used a small display and users stood much closer to the display, which may have influenced participants to behave more like they were interacting with a touch screen. Even in the other two studies, where participants were positioned further from the display, we saw them attempt to touch the screen. The changes to proximity and screen size were made to allow for easier collection of galvanic skin response data and to represent the scenario of using a kiosk to browse through a catalogue of furniture. This change for Study 2 could have easily reduced the magnitude of gestures being performed, as the objects on screen were much smaller, and our other studies showed that the size of the objects influences the speed and magnitude of the gestures (see Theme 3: Gestures for Size and Speed) (Morris et al., 2014).

To better understand whether these changes had any effect on the magnitude of the gestures produced, we looked at the body part used across studies (see Table 17).

Table 17

Percentage of gesture primitives for each part of the body for all 3 studies

Body part used	**Study 1: % of primitives using body part**	**Study 2: % of primitives using body part**	**Study 3: % of primitives using body part**
Arms (including hands, elbows, forearms, fingers)	60.80%	78.36%	68.33%
Legs (including feet, toes, knees)	7.49%	12.22%	4.55%
Full Body	22.92%	3.21%	9.70%
Head	4.20%	3.01%	3.33%
Eyes (gaze, blinking)	3.09%	1.80%	8.33%
Voice	1.51%	1.40%	5.30%
BMI	N/A	N/A	0.45%

Note. Highlighted cells indicate an increased production of that type of gesture compared to the other two studies.

As can be seen in Table 17, Study 1 has a much larger number of full-body interactions, which may have been influenced by either the lack of concreteness in the task, or might be evidence for priming. As we discussed previously, there is no consensus on the effect of priming in gesture elicitation studies yet (Cafaro et al., 2018; Hoff et al., 2016). We also see a much higher use of arms and legs and a much lower use of full-body movement for Study 2, which used the smaller display. Study 2 also had a lower use of voice, even though the study was conducted after Study 3 and may have been influenced by proximity to the screen. Finally, we see a significant jump in the use of voice and eye (gaze, blinking) interactions for Study 3 and also the introduction of brain-machine interaction technology, which are most likely due to changes in technology proliferation. See section 6.2.1 for a brief discussion on the proliferation of voice interaction technology.

When looking at the number of gesture primitives identified across studies, we see a significant reduction in primitives from Study 1 to Study 3. Study 1 had 133 primitives defined across 15 participants, but only 37.6% (n=50) of those gestures were mentioned by 2 or more participants. Many of these gestures were influenced by the priming, either because they were similar to motions shown in the video or performed during the kinesthetic priming (e.g. “doggy paddle”, “lunge”, “fishing”, and flying like an “airplane”). Study 2, which did not place an emphasis on the coding scheme or users’ mental models, had 72 unique gesture primitives. Finally, in Study 3, where the focus was on refining the qualitative coding scheme and many more features were introduced, 38 unique gesture primitives were identified. Some of the reduction in primitives was due to the fact that additional features that coded a subset of the movement, eliminating the need for added primitives. Another reason for the reduction in primitives could have been due to the lack of priming and the addition of concrete referents and concrete scenarios. In either case, we saw a significant decrease in the number of gesture primitives that were only specified by one participant and a reduction in variability (increase in consensus-distinct scores) between Study 2 and 3 where consensus-distinct scores were calculated. Study 2 consensus-distinct scores ranged from 0.421 to 0.556 (Table 6), while Study 3 scores ranged from 0.400 to 0.882 (Table 13).

One final difference across the studies was the role that hand-dominance played in user preference of the side of the body to use. In Study 1, we saw no real effect of hand-dominance on the side of the body selected with which to perform the gesture. What mattered more was the location of the object users were interacting with, reducing the need to reach across their bodies, which is consistent with research that shows that putting items that are often interacted with closer to the arm being used will also reduce fatigue (J. D. Hincapié-Ramos, Guo, Moghadasian, et al., 2014; Kölsch et al., 2003). In Study 3, however we saw a significant preference for the use of the right side of the body, and right-handed movements were found to be much faster (Nguyen & Kipp, 2014) and unlike for Study 1, all participants in Study 3 were right-handed.

6.1.1 The Rise of Voice Interactions

Over the course of the three studies, we see some interesting trends in voice interaction. While we do not see a significant difference between the use of voice between studies 1 and 2, there are other factors that could have influenced this outcome, such as the screen size or the types of actions presented in the referents. However, when comparing the two studies that were run with similar environments, we see a significant increase in the number of times voice was mentioned.

Figure 16 shows the launch of some of the major voice assistants on the market. While Siri launched in October 2011, our study conducted in mid-2012 did not see a significant influence of voice interaction. However, by the time the 2014 study was conducted, the technology had been out for 3 years, and two other voice assistants had launched. This trend toward ubiquity onto the market is likely the cause of the increase in preference for voice interaction.

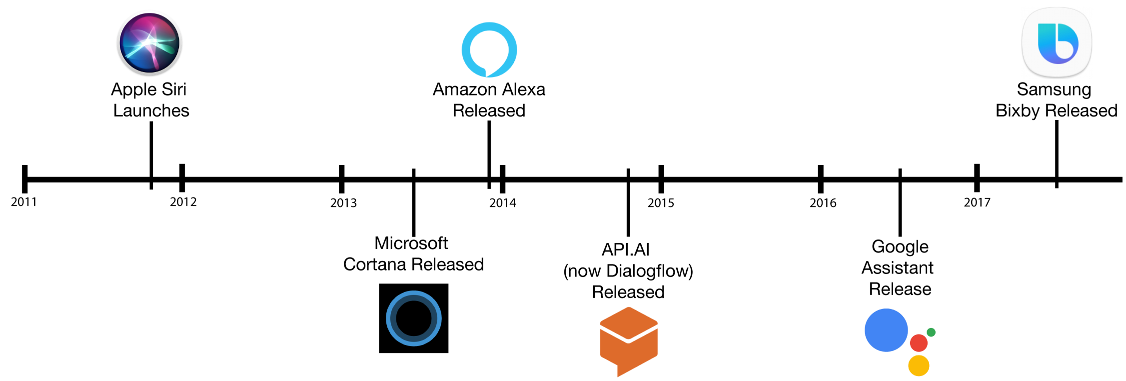


Figure 16. Timeline showing the launch of major voice assistants and technologies.

6.2 Emerging Themes Across Elicitation Studies

In Study 1, we introduced 6 themes, some of which we provided additional evidence for in Study 3. The 6 themes were:

- Theme 1: Similarities across Selecting Scrolling, and Paging
- Theme 2: Gestures on the Body, Gestures in Space
- Theme 3: Gestures for Size and Speed
- Theme 4: Gestures for Parallelizing Data Exploration
- Theme 5: Objects vs. Text
- Theme 6: Concerns about Ambiguity

From observations within these themes and additional results across all three studies, we can define the following design principles:

1) **Do not design gestures that require the user to maintain a specific palm direction.** We saw in Study 1 that palm direction was relatively fluid within and across participants. In Study 3, we saw more consistency, however this was also one of the features that was most likely to change when refining a gesture.

2) **Consider mapping interactions with large displays to the floor, the user's body or a smaller horizontal plane in front of the user**. In Study 1, we saw participants mapping the screen to their body or the floor. In Study 3, we additionally saw participants mapping the screen to a horizontal plane directly in front of them in mid-air like they were interacting with a touch table. We did not see these types of gestures in Study 2, where users were closer to the display and the display was small.

3) **Expect larger, less precise gestures for large objects or many objects.** Participants defined less precise and larger gestures when there are many objects or objects are large, which are more likely to exist when designing for large displays that are further away. In Study 2, we observed much smaller gestures as the screen was small and close to user sand the objects were small.

4) **Expect faster gestures when users want to navigate faster.** We saw participants comment on this in Study 3 and we also saw this for the *scroll* referent in Study 3, where there was a tie for max-consensus between slide (slower) and swipe (faster), but no such tie existed for *scroll fast*.

5) **Use more precise gestures drawing from metaphors of interactions with books or physical paper when users can interact with text instead of buttons or images.** Text not only required more precise gestures, but participants also drew off of different metaphors related to real-world physical interactions with paper and pencil. Example gestures for text including underlining, circling, or pulling instead of pushing text to a position in space.

6) **Design for a variety of joint rotations, magnitude of gestures, and variability in gesture path.** This principle is related to design principle 3, but here we go further. We saw that the primary joint of rotation changes and movements become smaller over the course of an interaction and a refinement, so designing a system that will recognize a swipe whether it is performed with a full arm gesture or just the hand will make the system more robust and make gestures easier to discover. Mapping straight or arching paths to the same movement will also make systems more robust.
7) **Don't design for discoverability, design for user preference and for minimizing fatigue.** We go further into this principle in the next section, 6.2.1.
8) **Design with multi-modal interaction in mind.** As we saw, voice, gaze, and even brain-machine interfaces are becoming more common for participants. Voice interaction can be leveraged to make interactions more accessible and can be used in instances where gestures may be considered ambiguous by users.

6.2.1 The (Un)Importance of Discoverability

So much of the gesture elicitation research, as mentioned previously, is focused on finding guessable gestures (Grijincu et al., 2014; Vatavu, 2012; Wobbrock et al., 2005, 2009) or on improving discoverability for gesture sets, especially if a user discovers one gesture and still needs to discover the rest (Cafaro et al., 2014, 2018). However, in this research we see across multiple studies, that what most elicitation studies would consider discoverable or guessable gestures are not necessarily the ones that are best suited for full-body gesture interaction due to legacy biases. The tension between discoverability and fatigue, which heavily influences user preference, is one that calls into question the importance of discoverability.

Cafaro et. al's approach is interesting because its primary focus is on providing allegories to frame the interaction such that discovering one gesture will lead to more easily understanding and discovering the remaining gesture set (Cafaro et al., 2014, 2018), which could be beneficial even if the emphasis is not on discoverability to begin with. Combining this approach, with an approach that minimizes legacy bias and gestural fatigue could be a promising area for future research.

So, if discoverability is not the primary goal, the problem becomes how do we quickly and easily guide users into understanding which gestures will work for public display interaction. One approach is that presented by Maher and Lee in the Walk-Up-and-Use Information Display (Maher & Lee, 2017), in which they provide short skeleton animations and visual guides to quickly train users and improve discoverability of their gesture set. Ultimately, the gesture set is part of a larger application that should have a well-designed feedback loop between system and user (see *Figure 17*) and provide the right affordances (Gibson, 1966; Norman, 2013) to users to more easily discover the gesture set.

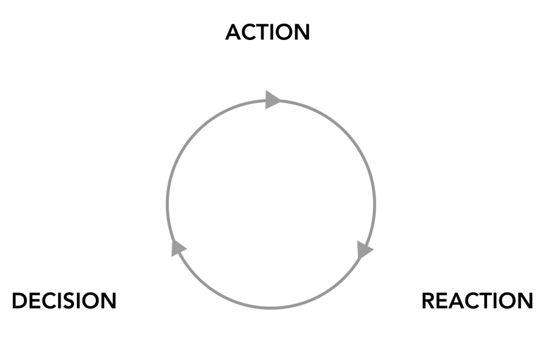


Figure 17. Feedback loop between system and user.

6.3 Implications for Gesture Elicitation

In the previous two sections, we highlighted the differences and similarities across the three elicitation studies conducted as part of this dissertation. These sections highlight some implications for gesture elicitation studies as whole.

First, as section 6.1 shows, the gestures that are elicited through this methodology are affected by various factors, such as the concreteness of the referents, the environment (private vs. public, the amount of physical space the can move around in, the display size), the diversity of participants (age, cultural background), and *legacy bias,* which changes over time as technology evolves. In this work, we explored some of these factors, although a more structured exploration is necessary. We also showed that increasing production helps combat legacy bias. Second, qualitatively coding gesture elicitation studies is not standardized and many papers do not present in detail the motivation for selecting the features that are selected for qualitatively coding gestures, making comparisons across studies difficult. In this work, we aimed to better understand which features are important to users' mental models in order to help guide this feature select in future studies. Third, calculating agreement does not guarantee a non-conflicting gesture set, and often one must look further that the most agreed upon gesture. In addition, as we presented in this work, discoverability and fatigue are often at odds with one another, and agreement scores focus largely on discoverability. Therefore, a more comprehensive and standardized approach to identifying a gesture set for a particular use case is still necessary.

While gesture elicitation is a promising avenue of research, and has been shown to produce gesture sets that are more preferred than designer specified gestures, more work should be done on improving the methodology, from the way users are elicited, to how gestures are coded, to how the codes are analyzed.

6.4 Sample Size Recommendations for Gesture Elicitation

The first elicitation studies conducted set a precedent of using a sample size of 20 participants per study regardless of the number of referents or types of gestures elicited (Wobbrock et al., 2005, 2009). A survey of 18 elicitation studies conducted since Wobbrock et al.'s 2005 study, excluding (Wobbrock et al., 2009) shows that these studies are conducted with a sample size between 8 and 31(M = 20.17, SD = 5.82) per condition (for studies that had multiple conditions the sample size was adjusted accordingly) , However, there is no clear indication for why a sample size of 20 participants is recommended or used, and no analysis has been conducted to determine whether there is a significant difference in a gesture set defined by gestures elicited from 10 participants or 20. Since gesture elicitation is a participatory design study that is analyzed using qualitative methods, we begin by looking at recommendations for qualitative studies more broadly.

For quantitative studies, there are clear guidelines for calculating statistical power and effect size, and therefore sample size (Cohen, 1988; Ellis, 2010). However, no clear guidelines exist for qualitative studies, and recommendations vary based on the type of study conducted, though they usually require a smaller sample size than quantitative methods.

Evaluation methods, such as usability studies, have low sample size recommendations since the goal of these studies is to identify pain points by participants of an existing interface. Nielsen & Landauer found that, for most interfaces, 100% of problems can be found by running a study with 15 participants (J. Nielsen & Landauer, 1993). Furthermore, Jakob Nielsen recommends small-N usability testing as 85% of all usability problems can be found by running only 5 participants in a study. One exception to this recommendation is for cases where there are multiple distinct groups that are being targeted by an interface, then one should run 5 participants per group. (J. Nielsen, 2010). This recommendation was confirmed by analyzing insights gained across 83 different studies (J. Nielsen, 2012).

Qualitative studies that use generative methods have different recommendations, as the goal of these methods is to better understand users' mental models, where there is more variability (J. Nielsen, 2004). For both ethnographic studies and grounded theory, Morse suggests running studies with 30-50 participants (Morse, 1994). However, for grounded theory, Creswell suggests only 20-30 (Creswell, 2007). For phenomenological studies, the recommendation from Creswell is 5-25 participants (Creswell, 2007), whereas Morse suggests no less than 6 (Morse, 1994).

For card sorting, which is a participatory design study aimed at understanding users' mental models, usually around information architecture, Nielsen recommends using 15 participants (J. Nielsen, 2004, 2012). Nielsen's recommendation is based off of Tullis & Wood's study, in which they originally tested 168 participants, and then simulated the outcome of the card sorting task with fewer participants and calculated how well the similarity scores correlate with the scores derived from testing the larger group (Tullis & Wood, 2004). 15 participants achieved a .90 correlation, while correlation of .95 was achieved for 30 participants. While Tullis & Wood recommended running 20-30 participants, Nielsen argues that the difference between .90 correlation and .95 correlation is not worth the added cost of running double the participants (J. Nielsen, 2004).

In many of the above cases, the sample size recommendations are built off of the notion of saturation, which occurs when adding more participants doesn't yield additional perspectives or data (Glaser & Strauss, 1967). More specifically, data saturation occurs when one can replicate the study with the existing amount of data and when adding more participants does not yield new codes, themes, or data (Fusch & Ness, 2015; Guest, Bunce, & Johnson, 2006; O'Reilly & Parker, 2012).

In this dissertation, we've shown repeatedly across the three studies that we were able to replicate many of the results for factors that remained static across elicitation studies, even with a small n of 10. For example, we saw evidence of the original themes that emerged in Study 1 in Study 3, in all studies we were able to reproduce the impact of production, and we saw similarities across the body parts used between Study 1 and Study 3 as well, Since there were some factors that changed across studies still, to confirm the ability to replicate the study another similar study could be run identical to Study 3, but this is left for future work. Additionally, in looking at the data in Study 3, we see that the data is layered, intricate, detailed, nuanced, etc., indicating a rich data set (Fusch & Ness, 2015).

To better understand the impact of adding more participants to an elicitation study, we look more closely at Study 3, in which many of the original shortcomings around qualitative coding were addressed. Each participant elicited an average of 17 unique gestures across all referents (S.D. = 3.53, min = 10, max = 22). *Figure 18* shows the total number of gesture primitives specified as each participant is coded. As seen in the figure, the majority of new gesture primitives are added in the first 3 participants, with very few new gestures specified after participant 5. Many of the new gestures specified by participants 4 and 5 do not have large agreement, with the majority of those gestures being specified by just one or two participants total. If running a larger sample of participants and only qualitatively coding a subset of them, then selecting the participants with the most gesture primitives specified will give you the largest variety of gestures (breadth).

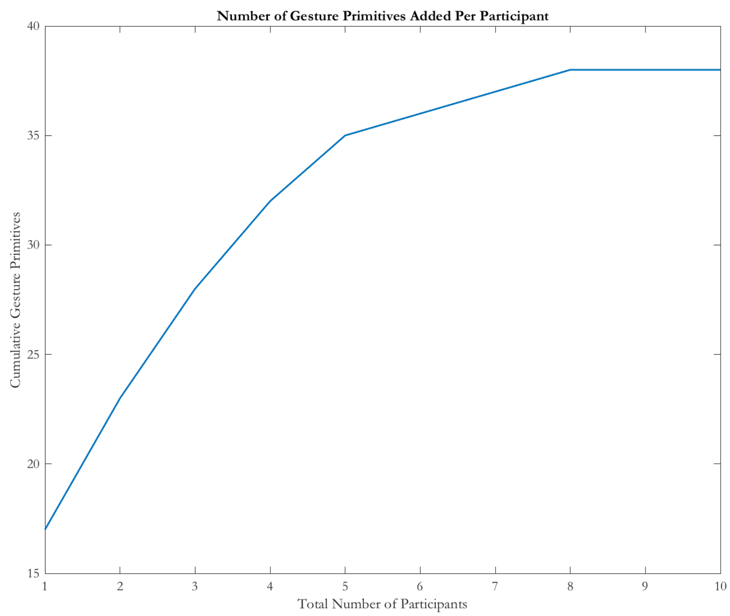


Figure 18. Cumulative gestures added for each participant coded in Study 3.

Since we see that very few new gestures are added after 5 participants have been coded, the next step was to analyze subsample permutations of the participants for n >=5 and see how closely the max-consensus (agreement) and consensus-distinct (variety) results match up to our n = 10 results. We ran 5 samples each of sizes n = 5, 6, 7, 8, 9 similar to how Tullis & Wood calculated the number of users needed for a card-sorting study (Tullis & Wood, 2004). The results can be seen in *Figure 19* for the max-consensus scores and *Figure 20* for the consensus-distinct scores. Table 18 shows a for a random sub-sample of the max-consensus scores, consensus-distinct scores and top three gestures for n = 5 to 9, with n=10 for reference.

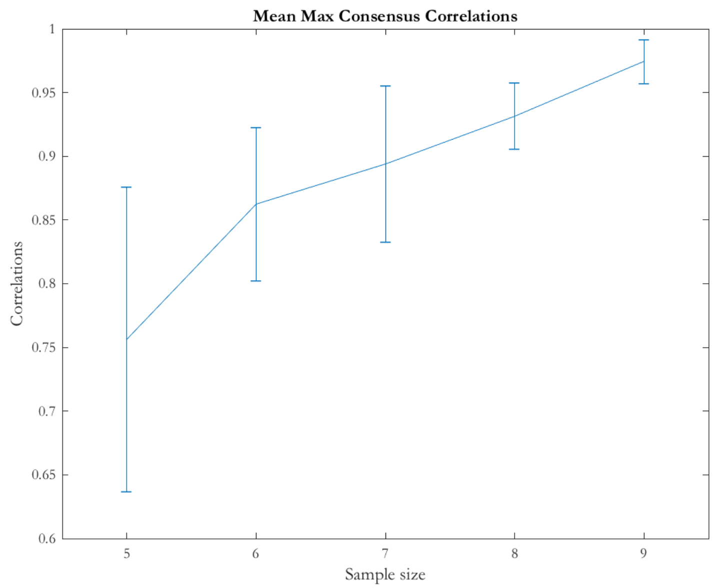


Figure 19. Means and standard deviations for the correlations between the max-consensus scores with n = 10 participants of 5 random samples of n=5, 6, 7, 8, 9.

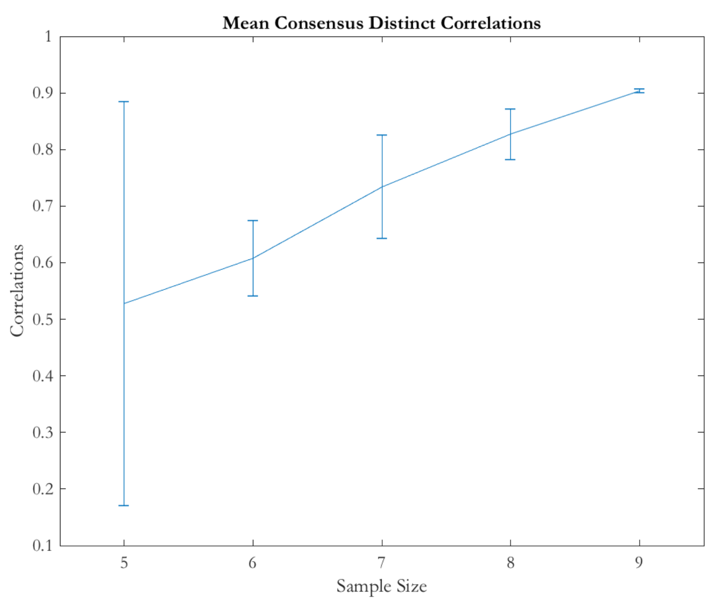


Figure 20. Means and standard deviations for the correlations between consensus-distinct scores with n = 10 participants of 5 random samples of n=5, 6, 7, 8, 9.

Table 18

Random sub-sample of the max-consensus, consensus-distinct and the three most agreed upon gestures for $n = 5$ to $n = 9$ participants.

n	5					6				
referent	**gesture**	**gesture 2**	**gesture 3**	**max-consensus**	**consensus-distinct**	**gesture**	**gesture 2**	**gesture 3**	**max-consensus**	**consensus-distinct**
'deselect'	tap	swipe	voice	80%	0.36	swipe	tap	voice	67%	0.22
'drag'	move	tap	pinch	100%	0.54	move	tap	voice	100%	0.71
'multi_object_select'	point	tap	voice	80%	0.50	point	voice	gaze	83%	0.50
'page_right'	swipe	step	voice	100%	0.42	swipe	step	voice	100%	0.42
'scroll'	swipe	step	voice	80%	0.55	swipe	voice	step	83%	0.58
'scroll_fast'	swipe	voice	gaze	100%	0.46	swipe	gaze	nod	100%	0.46
'select'	point	tap	voice	80%	0.33	point	tap	voice	83%	0.38
'zoom_in'	expand	voice	move	100%	0.28	expand	voice	tap	100%	0.28
'zoom_out'	pinch	tap	voice	60%	0.50	pinch	voice	tap	67%	0.58
n	**7**					**8**				
referent	**gesture**	**gesture 2**	**gesture 3**	**max-consensus**	**consensus-distinct**	**gesture**	**gesture 2**	**gesture 3**	**max-consensus**	**consensus-distinct**
'deselect'	swipe	tap	voice	57%	0.32	swipe	voice	tap	63%	0.40
'drag'	move	expand	tap	100%	0.79	move	voice	expand	100%	0.81
'multi_object_select'	point	gaze	voice	86%	0.41	point	gaze	voice	88%	0.41
'page_right'	swipe	step	voice	100%	0.54	swipe	step	voice	100%	0.47
'scroll'	swipe	voice	nod	86%	0.67	swipe	nod	rotate	75%	0.69
'scroll_fast'	swipe	nod	gaze	100%	0.62	swipe	gaze	nod	100%	0.53
'select'	point	tap	voice	86%	0.53	point	tap	voice	88%	0.67
'zoom_in'	expand	voice	tap	100%	0.50	expand	voice	tap	100%	0.61
'zoom_out'	pinch	voice	rotate	71%	0.67	pinch	voice	push	75%	0.53
n	**9**					**10**				
referent	**gesture**	**gesture 2**	**gesture 3**	**max-consensus**	**consensus-distinct**	**gesture**	**gesture 2**	**gesture 3**	**max-consensus**	**consensus-distinct**
'deselect'	swipe	tap	voice	56%	0.43	swipe	tap	voice	60%	0.42
'drag'	move	expand	grab	100%	0.76	move	grab	point	90%	0.82
'multi_object_select'	point	gaze	voice	89%	0.53	point	voice	gaze	90%	0.53
'page_right'	swipe	step	voice	100%	0.44	swipe	step	voice	100%	0.42
'scroll'	swipe	nod	slide	67%	0.69	slide	swipe	nod	60%	0.60
'scroll_fast'	swipe	nod	gaze	100%	0.60	swipe	nod	gaze	90%	0.71
'select'	point	tap	voice	89%	0.71	point	tap	voice	90%	0.67
'zoom_in'	expand	tap	voice	100%	0.63	expand	voice	point	100%	0.68
'zoom_out'	pinch	voice	tap	78%	0.50	pinch	voice	rotate	80%	0.50

Note. The same data is also shown for all 10 participants for reference. Highlighted cells indicate cases in which the gesture primitive is different from the next smallest random sample.

As can be seen in *Figure 19* and *Figure 20*, the max-consensus scores have a fairly low correlation at n = 5 and high standard deviation to the output of n =10 but quickly begin to converge. This makes it more likely that we see a significant amount of shifting around of gestures, as users are split on their preference for these.

Looking at the differences for one random sampling of n = 5, 6, 7, 8 and 9 and comparing it to n = 10 in Table 18, we see that very few of the most common gesture primitives (those that have the highest consensus scores) change after n = 6. However, the second and third most common gestures do still change.

Overall, based on the data presented, we can see that running 3 participants is enough to identify the majority of gesture primitives that users will specify and that running and analyzing 5-6 participants is enough to identify the most common gesture primitives across participants. However, analyzing additional participants will start to provide insights into gestures that are less influenced by legacy bias and into long-tail gestures that will have lower agreement scores. When trying to move past legacy bias, we need to run at least 10 participants, if not more depending on the depth of the analysis being conducted. 10 participants still provide a great amount of detailed information, as can be seen by the results and discussion presented in Chapter 5.

6.5 Limitations

This dissertation presents several studies aimed at gaining an in-depth understanding of user preference for gestural interaction, however the majority of the research focuses on able-bodied adults and does not account for additional factors, such as culture, that may influence user preference and mental models. In this section, we provide an overview of participant demographics that are not accounted for in this research, as well as highlight remaining open questions and potential future work.

6.5.1 Designing for Accessibility

Designing for accessibility means designing for "any user, anywhere, anytime" (Soegaard, 2019). As mentioned, in this research, our specific focused was on able-bodied adults, however public displays should be available to all users, many of which are not accounted for in the research presented here. In this section, we look at a few of the demographics that may struggle with gestural interactions with a full-body walk-up-and-use public display.

Visually Impaired

Visually impaired users may struggle with seeing the items on the display to interact with them and may have even more difficulty identifying that the display is interactive, an issue even for sighted individuals. When it comes to reading items on screen, standard accessibility tools already exist for visual displays on desktop and mobile. These same screen-reading tools could be implemented on large public displays as well. The concern then becomes communicating to users how to trigger the screen-reading capabilities. One way, might be to also provide voice control capabilities, a feature that was regularly suggested by participants in elicitation studies. Additionally, computer vision capabilities inherent in gestural interfaces could allow for recognition of support or probing canes, allowing the system to automatically switch modalities and engage a user via voice.

Since gestures can still be performed by visually impaired users, studies have been conducted both to aid visually impaired users in understanding other people's gestures in collaborative environments and in identifying differences in gesture preferences between visually impaired and sighted individuals. To help visually impaired users in collaborative environments, Kunz et al. developed a system that processed gestures and speech in order to provide meaningful information about deictic gestures performed by others in tabletop interactions (Kunz et al., 2014).

To better understand user preference of visually impaired individuals, Kane et al. conducted a gesture elicitation study for touch-based gestural interaction comparing gestural preferences for both blind and sighted individuals, and found that these two groups prefer different types of gestures (Kane, Wobbrock, & Ladner, 2011). During the elicitation study, blind participants suggested gestures that used the edges or corners of the screens, contained significantly more strokes, and used more multi-touch gestures, many of which used a mode key while performing a different gesture with the other hand. Additionally, more of the gestures produced by blind participants were abstract or metaphorical, while more of the gestures produced by sighted people were symbolic. For metaphorical gestures, blind participants often used references to physical keyboards, whereas none of the sighted participants did, which points to a different type of legacy bias than that observed in our study and similar studies with sighted participants. Blind participants also performed larger gestures, gestures had a greater variation in size across instances, glyph gestures, specifically, were wider, and all gestures took twice as long to perform as it took sighted users (Kane et al., 2011), The difference in how long it takes individuals to perform gestures means that speed, although a meaningful feature for sighted participants, is one that cannot be used if an interface is to be accessible for all users. Other demographics, such as older individuals, are also slower in performing gestures (see Elderly section below). This study was conducted for touch interactions, where edges and corners exist and can be felt. However, in full-body gesture interaction, the camera frame is difficult to communicate to users, and cannot be felt or touched. However, audio feedback could potentially be used to indicate when users are approaching the bounding box of the recognition space.

For mid-air gestural interaction, Funes et al. conducted a study evaluating the use of gestural interaction for video playback to help visually impaired users through a system called Gesture4All. They tested several gesture recognition technologies, and found that 3D gesture interaction using a smart phone was preferred over standard accessibility tools, such as screen reading software, for video playback because it was easier to perform accurately and the Gesture4All system also provided audio feedback when a gesture was recognized (Funes, Fortes, Trojahn, & Goularte, 2018). This study provides additional support for the user of multimodal output for users, and suggests a novel approach for gesture interaction. An alternative for visually impaired users, could be to use tactile feedback and the accelerometer on their own mobile devices to allow for communication with a public display via blue tooth pairing, for example.

Physical Impaired

There are many different types of physical impairment that one might be concerned about. Much of the research has focused on individuals with motor impairments, such as degenerative diseases or others that cause tremors. However, individuals may also be wheelchair bound or be missing limbs. Research into individuals with these types of impairments is much less common.

For users with motor impairments, several researchers have looked at how to modify or augment touch-based interactions, as these screens are small and therefore require precise movements that some are not able to perform. For example, in stylus-based touch interaction for small screens, Wobbrock et al. designed and evaluated a unistroke input technique that uses corners and edges as guides. This system increased accuracy for able-bodied individuals by 18% and could also be used by individuals with motor-impairments, such as tremors, even when those users were not able to use the Graffiti writing system (Wobbrock, Myers, & Kembel, 2003). Malu et al. evaluated the use of smartwatch interactions with users that have motor impairments and found that text input is extremely difficult without voice, and that even though users thought tap interactions were easy to perform, they had a high error rate. In their study, the authors also conducted an elicitation study for smart watch interaction by users that are motor impaired. In their second study, they observed legacy bias, but also noticed that legacy bias didn't have as strong of an effect with this demographic, where there are limitations in motion and comfort is even more important. Additionally, they found that users that have motor impairments of the upper limbs explicitly disliked on-body interactions, for smart watches (Malu, Chundury, & Findlater, 2018). In the first study presented in this dissertation, we saw a trend for users to elicit on-body gestures with large displays. This difference could be due to either the significant change in screen size, or it could highlight a difference between able-bodied and motor impaired individuals.

In full-body gesture interaction, one elicitation study, conducted specifically looked at the effect of physical impairment in gesture elicitation studies (Altakrouri, Burmeister, Boldt, & Schrader, 2016). The study was conducted with 20 healthy and 12 physically impaired participants in an office setting and found that impaired participants were more likely to use full body motions, had lower agreement scores, and wanted more personalized gestures. An interesting finding in this particular study was that users specified that they would not use full body movements in public, but they would in private. This could be due to cultural influences, as the study was conducted in Germany, or could specifically pertain to a public vs. private office setup and not to all public spaces. In the first study presented in this research, we found a significant number of full body interactions suggested that users would be willing to perform in public, especially around data navigation tasks. Another differentiation between the study and the research presented here is that this research explicitly looks at walk-up-and-use displays, where personalization is not possible. Additional research would need to be conducted to better understand how motor impairment affects users of different cultural backgrounds and how user preferences are affected in different contexts.

Physically impaired participants can benefit from gaze and voice, both of which were mentioned by participants in our elicitation studies. Some participants were explicitly concerned about the perceived ambiguity of gaze as an interaction modality, however research has shown that this can be an effective mode of interaction (Chakraborty, Sarcar, & Samanta, 2014; Jacob, 1991; Majaranta, Aoki, Donegan, Hansen, & Hansen, 2011; Rajanna & Hammond, 2018; Wobbrock, Rubinstein, Sawyer, & Duchowski, 2008) for text entry (Chakraborty et al., 2014; Wobbrock et al., 2008) as well as actions such as minimizing and maximizing windows, scrolling, refreshing a website or opening a new tab (Rajanna & Hammond, 2018). During our elicitation studies, sometimes participants would consider individuals that could not walk or move around, such as those in wheelchairs and would suggest small gestures that could be performed while sitting (e.g. drawing something or performing a gesture on the armrest of the couch).

Elderly

While much of the research on motor impairment and visual impairment may also be applicable to the elderly, older adults also undergo sensory, perceptual and cognitive changes (Schieber, 2003) and declining motor skills (Vercruyssen, 1997) as they age. This may impact their ability to perform full-body gestural interactions. They may also suffer from conditions, such as strokes, which may impair only one side of their body. In the studies presented in this dissertation, all participants were adults between 18 – 48 years of age, and unlike situational impairments or motor impairments, most of our participants did not comment on age related modifications that could be made to gestures as they were elicited.

Several researchers have evaluated older adult (either over the age of 60 or 65 depending on the study) usage of touch screen interfaces. One study comparing younger and older adults on touch gesture interaction found no difference in accuracy between the two groups regardless of gesture complexity or screen size, but found that older adults were often slower in performing the gestures (Stößel, Wandke, & Blessing, 2010). Another study comparing older and younger participant's preference for touch gestures found that older adults prefer one-finger over multi-finger gestures (e.g. double tap to zoom instead of pinch-to-zoom), are more tolerant of more complex gestures, and prefer symbolic gestures more often than younger users (Stößel & Blessing, 2010). When looking at these results, this may also mean that older individuals are less likely to be influenced by legacy bias of touch screen interfaces. In yet another study by Kobayashi et al., the researchers found that older adults had difficult with tapping accuracy, especially with small targets and that interactions with large screens often outperformed small screens (Kobayashi, Hiyama, & Miura, 2011).

One advantage of large, public displays is that objects on screen tend to be larger and easier to see and most full-body gesture recognition technology allows users to be closer or further away as needed, allowing older adults to approach the screen if they are having difficult reading it. In our studies, we also found hand-dominance did not often affect with which side of the body that users chose to perform a gesture, indicating that many gestures could be performed with either the right or left side of the body and contain the same meaning. Allowing user to, for example, swipe with either the right or left hand and focus on direction of motion instead, would allow elderly individuals that have control over only one side of their body to still interact with displays. Arm gestures, across all studies, were also most common. Designing public displays with only upper body interactions would also provide a more accessible interface for physically impaired individuals and elderly users that are wheelchair bound. Again, the speed at which a gesture is performed, cannot be used as a distinguishing feature due to the differences in speed between younger and older individuals.

For mid-air gestural interaction, one study, aimed at comparing older adults with younger ones, found that older individuals are more likely to rate gestures as fatiguing and difficult to perform, were more affected by failed attempts, and found it harder to perform gestures that required their hand to be in a specific location than younger users (Cabreira & Hwang, 2016), providing evidence for preferring to design a system with gestures that are relative to one's own body instead of in a particular point in space. Cabriera & Hwang also conducted two additional studies, one to assess swipe gestures (Cabreira & Hwang, 2018b) and another to assess pointing and selection gestures (Cabreira & Hwang, 2018a). For swipe they found that larger carousel menus encourage older adults to make larger movements, thereby increasing recognition rates (Cabreira & Hwang, 2018b). This is consistent with our findings that larger objects elicit larger movements from participants. For pointing and selection tasks, they found that providing both audio and visual feedback helps older adults increases target acquisition speed but not accuracy and that target location significantly affects both selection time and accuracy, and that targets in the top right or towards the center of the screen were most successful (Cabreira & Hwang, 2018a). Another study aimed to develop a full-body gesture-based game for institutionalized older adults. In this research, the authors worked with a physical therapist to develop a gesture set that would be accessible for older adults. In their evaluation, they found that gestures that required leg movements or both arms were rated as more difficult, but that participants liked that the gestures allowed them to be active, with some participants even rating the gestures as too easy. Overall, dynamic gestures were less likely to be completed than static ones (Gerling, Livingston, Nacke, & Mandryk, 2012). This study builds on existing research to show, that even though the elderly may struggle with aspects of gestural interfaces, they are still motivated to engage with them.

Children

While children are less likely to suffer from physical impairments, such as tremors, or be wheelchair bound, children are still developing their motor control, especially fine motor movement. The lack of fine motor control is especially important for touch interaction, where there is limited screen real estate and movements are much smaller. Therefore, the majority of research into children and gesture has been done for touch, especially since children as young as age 2 are using tablets and other touch interfaces on their own.

When evaluating the seven most common gestures for touch interactions (tap, flick, drag-and-drop, slide, pinch, spread and rotate), Aziz found that children aged 4 and older could perform all gestures with no difficulty, while children aged 3 struggled with pinch and spread gestures, and children aged 2 additionally struggled with rotate and drag-and-drop (Aziz, 2013). Similarly, Nacher et al. evaluated 100 apps on the app store for iOS and found that nearly all of them only used tap and drag. However, when evaluating what gestures children aged 2-3 could perform, they found that these children could also perform one-finger rotations and two-finger scale up / down gestures with the same accuracy as tap and drag. They found that children had difficulty performing double tap, long press and two-finger rotation gestures (Nacher, Jaen, Navarro, Catala, & González, 2015).

To understand whether children and adults elicit different gestures, Rust et al. conducted an elicitation study with children from 8 to 11 years of age and adults for touch interactions. Similar to other studies, they found that both children and adults are influenced by legacy bias (96% of gestures elicited were standard touch screen gestures, such as tap, drag, swipe, pinch and rotate), and that children and adults elicit similar gestures (Rust, Malu, Anthony, & Findlater, 2014). Similar to the second and third study presented in this dissertation, Rust et al. used concrete referents, but unlike standard elicitation methods, they provided feedback when the touch was registered; they also allowed children to use a drawing program for 5 minutes before the study to eliminate novelty of the technology, which may have biased the children in some way.

In another study, Anthony et al. compared gesture recognition accuracy for adults and children (7 – 16 years of age), finding that children are more likely to have holdover touches (touches located near the previous touch target) and higher rates of errors for touches in general. Additionally, children's gestures were less likely to be recognized than adult gestures and that recognition accuracy was correlated with age – the younger the child the worse the recognition (Anthony, Brown, Nias, Tate, & Mohan, 2012). In prior work by Brown et al., they also found that swipe gestures may be difficult for children because they tend to lift their finger mid swipe and that children have difficulty acutely touching small onscreen targets (Brown et al., 2010).

For full-body gestural interaction, Connell et al. conducted an elicitation study with six children (ages 3-8) asking children to produce gestures for 22 tasks using abstract shapes of 3 different categories (object manipulation, special interaction, and utility or navigation based tasks) (Connell, Kuo, Liu, & Piper, 2013). In their study, they found evidence, even in young children, for legacy bias, but found that younger children that had less experience with Kinect or touch interactions were more likely to suggest egocentric gestures, voice interaction, full-body gestures, and generally produced a larger variety of gestures. This study, however, had a very small sample size and used a Wizard-of-Oz approach providing feedback to participants during the elicitation process. Regardless, this study highlights that, young children with less or no touch interaction experience may be a promising avenue for further study, as they may provide better insights into natural gestures that are not influenced by legacy bias.

In general, however, there is little research into designing for full-bodied interactions for children, especially for task driven public displays as children are also less likely to use these over interactive museum displays or gesture-based games, either in public or at home. Concerns around small target sizes and precise movements are less likely to be problematic when dealing with large public displays, as targets can be larger, and non-finger movements do not require the level of precision that finger gestures for touch interaction require.

Situational Impairments

Every user undergoes situational impairments (Sears, Lin, Jacko, & Xiao, 2003) at some point in time, and users are especially likely to be experiencing situation impairment in locations where walk-up-and-use public displays would be found. When thinking about public displays in airports and malls, users may be carrying items they just purchased, ensuring their children do not run off, be stressed about making a flight, or other contextual circumstances that either lead to decreased cognitive ability or motor ability. One participant, in Study 3, specifically, thought about what it would be like to be holding her child in her arms, or holding groceries and still be able to interact with the display. This was one of the primary motivating factors for her when producing gestures.

Early work in situational impairment argues that environmental and contextual changes might affect mobile device users similar to how cognitive and physical impairments affect users with disabilities (Barnard, Yi, Jacko, & Sears, 2007). One study found that the number of errors made by an unimpaired user on a mobile device was similar to motor-impaired desktop users (Yesilada, Harper, Chen, & Trewin, 2010). Similarly, another study looking at motor impairment, also found perceived benefits to using smartwatches with able-bodied situationally impaired individuals. (Malu et al., 2018).

Some research has been conducted in touch and desktop-based interaction to help identify and support situational impairment. For example, many people use their phones while walking. One study, looked at improving typing accuracy by using the phone accelerometer to help compensate for the user's movement when walking (Goel, Findlater, & Wobbrock, 2012). Another study, by Rajanna and Hammond looking at desktop interactions, suggest using gaze interaction as a way to address situational impairment

(Rajanna & Hammond, 2018).

Some insights can be drawn from full-body gesture research into the elderly and physically impaired individuals, such as supporting flexibility in which arms can be used for common gestures (such as swipe, tap, etc.), and supporting multi-modal input and output. However, to the best of our knowledge, no research has been conducted on how to account for situational impairment specifically in full-body walk-up-and-use interactions, making it a rich area for future research.

6.5.2 Effect of Culture on Gestures

Cultural norms can also influence the types of gestures users are most likely to perform. This is especially important for walk-up-and-use interfaces, because users' diverse cultural backgrounds are likely to interact with them. Some HCI research has sought to better understand how culture influences gesture elicitation.

For example, researchers from Stanford conducted two similar elicitation studies: one in the US and another in China, to better understand how culture influences user preference for interacting with drones (Cauchard et al., 2015; E et al., 2017a). Both elicitation studies were conducted on university campuses and participants were presented 18 different actions associated with drone control (e.g. follow me, fly higher, take a picture). Participants were asked to interact in the manner that felt most natural (not specifically gesture). The study used a post-task think-aloud method and asked participants to rate suitability and simplicity. After the elicitation portion of the study (part one), participants were given a sheet of paper with suggestions for interaction techniques and asked to complete 4 of the 18 tasks again to see if their interaction changed after the suggestions (part two). US participants expressed discomfort using voice initially and use of voice increased while use of gesture decreased between part one and part two of the study (from 37% to 57% for voice and 88% to 70% for gesture). However, Chinese participants used voice more often from the very beginning of the study and there was no significant difference between modalities chosen in part one and part two of the study (56% and 59% respectively for voice and 80% to 84% for gesture). Chinese participants were also more likely to use multimodal commands. In 75% of cases voice input would complement the gesture, but in the remaining 25% the voice input would augment the gesture by providing additional information. Finally, although there was significant agreement within cultures and for many actions there was also significant agreement across cultures, there were instances in which there was high agreement in one culture and low agreement in another (e.g. with stop, where Chinese participants use two common gestures for stop – holding their hand out in a "T-shape" and holding their palm out, like in the US).

A significant amount of research has looked at the effect of culture on gesture outside of the HCI literature, and a full review of this work is outside the scope of this dissertation. However, a couple of examples will be discussed.

Matsumoto and Hwang compared emblematic gestures, those gestures that don't need to co-occur with speech, across cultures (Matsumoto & Hwang, 2013). They first had encoders produce emblems, and then they had an independent set of observers judge the emblems produced by the encoders to make sure they were really emblematic gestures. A list of gestures was derived from previous studies using the same encoding/decoding methodology in three different countries: the US, Israel and Iran. A subset of the most important ones, as chosen by three independent subject matter experts was chosen for the comparison study. In this study, they found that 15 of the gestures in their study were recognized across gestures. These gestures were often basic gestures (e.g. “no”, “yes”. “to threaten someone” by waving your fist at them, etc.) or gestures referring to objects that are culturally invariant (e.g. “run”, “phone” and “cigarettes”). They found no culturally similar gestures that were categorized as religious ones or ones referring to social norms and etiquette. Research conducted by Archer on cultural differences in gesture ultimately led the author to conclude that there is no "universal language" of gestures (Archer, 1997). This was supported by differences such as the meaning of “OK” or “Thumbs up” gestures in English having many different, often negative interpretations in other languages.

For walk-up-and-use public displays and elicitation studies in general, this means that researchers should try to both identify common gestures that might be insulting in other cultures prior to developing a final gesture set and to recruiting participants from diverse cultural backgrounds. Not only may gestures mean different things in different languages, but the social acceptance of performing gestures in public or using voice interaction in addition to gesture will differ. This may also mean that, just like mobile applications, commercial applications developed for multiple locales may need slightly differing interaction paradigms, not just text localization. As public displays and full-body or free-space gestural interaction becomes more common, additional research should be conducted to determine the impact of culture on developing systems that are deployed across cultures.

6.6 Open Questions and Future Work

While this dissertation highlights the importance of fatigue and comfort for participants and asks users to qualitatively assess the fatigue of gestures they perform, the research presented here shows that there is still no quantitative measure of fatigue that can be used for the variety of free-space and full-bodied gestures that are elicited for walk-up-and-use interactions. Additionally, while the variability of elicited gestures changed between studies, it is not clear whether this variability was affected by priming or by the fact that one study asked participants to interact with abstract objects and another with concrete objects. Both of these questions are interesting avenues for future work.

Another interesting avenue to explore is on how to leverage the methodological changes presented here to de-emphasize initial discoverability, and find gesture sets that users are prefer and which are less fatiguing. This avenue of research could take multiple approaches. One could be to explore how to quickly communicate to users what the gesture set is for a system. Another could be to leverage framed guessability (Cafaro et al., 2018) to more easily allow users to draw off of allegories to discover additional gestures once one has been discovered.

While a significant amount of research has been conducted in understanding the impact of culture on gestures in human-to-human communication, very little research exists around how culture influences elicitation of gestures or gestural preference in public displays. Similarly, there is research in how to design accessible touch interactions, but much less work in full-body or free-space gestures. In this dissertation, we presented design principles that could be used for full-body interactions with public displays for able-bodied adults, but public displays should be accessible to all demographics. Therefore, another avenue of research is to identify how the design principles presented here should be refined such that they are accessible for all demographics.

Finally, a logical next step for future work would be to train more robust classifiers for a walk-up-and-use system and assess how easily users can discover the gestures they need to perform without any training.

CHAPTER 7

CONCLUSION

The goal of the research presented in this dissertation was to answer the following research questions:

RQ 1) How do we modify gesture elicitation to reduce legacy bias?

RQ 2) Which gestural features matter to users and how do they influence a user's mental model about that gesture?

RQ 3) What are the set of design principles that can be used in the future to design gestural interfaces that are discoverable, easy-to-use-and flexible for public displays?

Answering these questions is especially important now given the rise of post-WIMP interaction technologies. Existing research has already shown that users prefer user-generated gestures to those produced by expert designers. Using a gesture elicitation methodology, therefore, is an ideal way to identify gestures set for walk-up-and-use free space contexts.

In this dissertation, we presented three gesture elicitation studies to address the research questions presented above. The first study focused on RQ1 and explored the use of priming (both kinesthetic and video priming) and increased production as ways to combat legacy bias. Priming did not seem to have a significant effect on the number of gestures elicited per participant. However, there were indications that priming did influence the types of gestures people performed (e.g. by mimicking gestures they did during the kinesthetic priming phase, or ones they saw in the video priming) and results from this study and related work suggest priming should be further explored. This study also indicated that increases in production were found to be beneficial, showing that users typically preferred the second or third gesture produced per referent. This was corroborated by the subsequent studies as well.

The second study focused on RQ3 by exploring the interplay between fatigue, discoverability and user preference. In this study, we found that users care considerably about fatigue and comfort when performing gestures. However, they were not good at judging how fatiguing a gesture was during the elicitation phase, and often changed their preferences after repetition. Finally, we found that discoverability and fatigue are at odds with one another, and that gestures that are more discoverable also tend to be more fatiguing and therefore less preferred by participants.

The third study focused on RQ2 to identify the features that are most important to users' mental models to both identify design principles that can be leveraged by designers and better inform the development of robust classifiers for full-body gesture based systems. We discussed changes in the qualitative coding to help identify and distinguish gestures that are considered similar and different by users. We also explored what refinements tell us about user preference, and found that refinements were often done for gestures that users already preferred and were refining exclusively arm gestures. We identified palm direction and hand configuration are largely variable and containing little meaning for participants and found that participants also used arching paths in addition to straight paths without any change in the meaning of the gesture. Finally, in this study, we also showed that current quantitative measures of fatigue, all of which focus on shoulder joint rotation, are inadequate for many of the gestures that are elicited from participants for full body walk-up-and-use interactions.

Across the three studies, we identified 8 design principles. The following is a summary of the design principles that emerged across the three studies for walk-up-and-use interactions:

1) Do not design gestures that require the user to maintain a specific palm direction.
2) Consider mapping interactions with large displays to the floor, the user's body or a smaller horizontal plane in front of the user.
3) Expect larger, less precise gestures for large objects or many objects.
4) Expect faster gestures when users want to navigate faster.
5) Use more precise gestures drawing from metaphors of interactions with books or physical paper when users can interact with text instead of buttons or images.
6) Design for a variety of joint rotations, magnitude of gestures, and variability in gesture path.
7) Don't design for discoverability, design for user preference and for minimizing fatigue.
8) Design with multi-modal interaction in mind.

The key contributions of this thesis are:

1) A modified gesture elicitation methodology that aims to overcome legacy biases through the use of priming and increased production.
2) A qualitative coding scheme that better captures users' mental models.

A set of generalizable design principles that can be used in the future to design gestural interfaces that are discoverable, easy-to-use and flexible for walk-up-and-use public displays.

REFERENCES

Acer. (2015). Acer to Bring Intel RealSense 3D Camera to Its Powerful Aspire V 17 Nitro Notebook PCs. Retrieved from https://www.acer.com/ac/en/US/press/2015/151402

Alexander, J., Han, T., Judd, W., Irani, P., & Subramanian, S. (2012). Putting Your Best Foot Forward: Investigating Real-World Mappings for Foot-based Gestures. In *CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1229–1238). Austin, Texas, USA: ACM Press.

Altakrouri, B., Burmeister, D., Boldt, D., & Schrader, A. (2016). Insights on the Impact of Physical Impairments in Full-Body Motion Gesture Elicitation Studies. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)* (pp. 5–15). New York, NY, USA: ACM Press.

Anthony, L., Brown, Q., Nias, J., Tate, B., & Mohan, S. (2012). Interaction and Recognition Challenges in Interpreting Children's Touch and Gesture Input on Mobile Devices. In *ITS '12 Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (pp. 225–234). Cambridge, Massachusetts, USA: ACM Press.

Archer, D. (1997). Unspoken Diversity: Cultural Differences in Gestures. *Qualitative Sociology*, *20*(1), 79–105.

Aziz, N. A. A. (2013). Children's Interaction with Tablet Applications: Gestures and Interface Design. *International Journal of Computer and Information Technology*, *02*(03), 447–450.

Bachynskyi, M., Palmas, G., Oulasvirta, A., Steimle, J., & Weinkauf, T. (2015). Performance and Ergonomics of Touch Surfaces: A Comparative Study Using Biomechanical Simulation. In *Proceedings of CHI 2015* (pp. 1817–1826). Seoul, Republic of Korea: ACM Press.

Baldwin, R. (2013). Asus and Leap Motion Bring Minority Report-Style Gestures to Life. Retrieved from https://www.wired.com/2013/01/leap-motion-asus/

Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2007). Capturing the effects of context on human performance in mobile computing systems. *Personal and Ubiquitous Computing*, *11*(2), 81–96. http://doi.org/10.1007/s00779-006-0063-x

Baudel, T., & Beaudouin-lafon, M. (1993). CHARADE : Remote Control of Objects using Free- Hand Gestures. *Communications of the ACM*, *36*(7), 28–35.

Blackshaw, M., DeVincenzi, A., Lakatos, D., Leithinger, D., & Ishii, H. (2011). Recompose: Direct and Gestural Interaction with an Actuated Surface. In *Proceedings of CHI 2011* (p. 1237). Vancouver, BC, Canada.

Borg, G. (1998). Human Kinetics. *Borg's Perceived Exertion and Pain Scales.*

http://doi.org/10.1123/pes.2016-0084

Boring, S., Jurmu, M., & Butz, A. (2009). Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. *21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OzCHI'09)*, 161–168. http://doi.org/10.1145/1738826.1738853

Brown, Q., Bonsignore, E., Hatley, L., Druin, A., Walsh, G., Foss, E., … Golub, E. (2010). Clear Panels: a technique to design mobile application interactivity. In *Proceedings of the Conference on Designing Interactive Systems* (pp. 360–363). Aarhus, Denmark: ACM Press. http://doi.org/10.1145/1858171.1858237

Cabral, M. C., Morimoto, C. H., & Zuffo, M. K. (2005). On the usability of gesture interfaces in virtual reality environments. In *Proceedings of the 2005 Latin American conference on Human-computer interaction - CLIHC '05*. http://doi.org/10.1145/1111360.1111370

Cabreira, A. T., & Hwang, F. (2016). How Do Novice Older Users Evaluate and Perform Mid-Air Gesture Interaction for the First Time? In *Proceedings of NordiCHI '16*. Gothenburg, Sweden: ACM Press.

Cabreira, A. T., & Hwang, F. (2018a). Evaluating the Effects of Feedback Type on Older Adults' Performance in Mid-Air Pointing and Target Selection. In *Proceedings of SUI '18*. Berlin, Germany: ACM Press.

Cabreira, A. T., & Hwang, F. (2018b). Movement Characteristics and Effects of GUI Design on How Older Adults Swipe in Mid-Air. In *Proceedings of ASSETS '18* (pp. 420–422). Galway, Ireland: ACM Press.

Cafaro, F. (2012). Using Embodied Allegories to Design Gesture Suites for Human-Data Interaction. In *Ubicomp* (pp. 560–563). Pittsburgh, PA, USA: ACM Press.

Cafaro, F., Lyons, L., & Antle, A. N. (2018). Framed Guessability : Improving the Discoverability of Gestures and Body Movements for Full - Body Interaction. In *Proceedings of CHI 2018* (pp. 1–12). Montreal, Quebec, Canada: ACM Press.

Cafaro, F., Lyons, L., Kang, R., Radinsky, J., Roberts, J., & Vogt, K. (2014). Framed Guessability: Using Embodied Allegories to Increase User Agreement on Gesture Sets. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 197–204). Munich, Germany: ACM Press. http://doi.org/10.1145/2540930.2540944

Cauchard, J. R., E, J. L., Zhai, K. Y., & Landay, J. A. (2015). Drone & Me: An Exploration Into Natural Human-Drone Interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. New York, NY, USA: ACM Press. http://doi.org/10.1145/2750858.2805823

Chakraborty, T., Sarcar, S., & Samanta, D. (2014). Design and evaluation of a dwell-free eye typing technique. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14* (pp. 1573–1578). Toronto, ON, Canada: ACM Press. http://doi.org/10.1145/2559206.2581265

Claude, C., Cerma, V., & Carbonell, N. (1993). AN EXPERIMENTAL STUDY OF FUTURE "NATURAL" MULTIMODAL HUMAN-COMPUTER. In *INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems* (pp. 67–68). ACM Press.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edi). New York, New York, USA: Lawrence Erlbaum Associates.

Connell, S., Kuo, P.-Y., Liu, L., & Piper, A. M. (2013). A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13* (pp. 277–280). New York, New York, USA: ACM Press. http://doi.org/10.1145/2485760.2485823

Creswell, J. W. (2007). *Qualitative Inquiry & Research Design: Choosing Among Five Approaches* (Second Edi). Thousand Oaks, CA: SAGE Publications. http://doi.org/10.1016/S0022-3476(89)80781-4

Csikszentmihalyi, M. (1997). *Finding Flow: The Psychology of Engagement with Everyday Life*. Basic Books Inc.

Dourish, P. (2001). *Where the action is: The foundations of embodied interaction*. MIT Press.

Dow, S. P., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D. L., & Klemmer, S. R. (2011). Prototyping Dynamics : Sharing Multiple Designs Improves Exploration , Group Rapport , and Results. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1–10). Vancouver, BC, Canada.

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., & Klemmer, S. R. (2010). Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction*, *17*(4), 1–24. http://doi.org/10.1145/1879831.1879836

Dow, S. P., Heddleston, K., & Klemmer, S. R. (2009). The efficacy of prototyping under time constraints. In *Proceeding of the seventh ACM conference on Creativity and cognition - C&C '09*. Berkeley, California, USA: ACM Press. http://doi.org/10.1145/1640233.1640260

E, J. L., E, I. L., Landay, J. A., & Cauchard, J. R. (2017a). Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. http://doi.org/10.1145/3025453.3025755

E, J. L., E, I. L., Landay, J. A., & Cauchard, J. R. (2017b). Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques. In *Proceedings of the 2017 CHI Conference on*

Human Factors in Computing Systems (CHI '17) (pp. 6794–6799). New York, NY, USA: ACM Press. http://doi.org/10.1145/3025453.3025755

Efron, D. (1941). *Gesture and Environment*. King's Crown Press.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes*. Cambridge, Massachusetts, USA: Cambridge University Press. http://doi.org/10.1017/CBO9780511761676

Epps, J., Lichman, S., & Wu, M. (2006). A study of hand shape use in tabletop gesture interaction. In *CHI '06 extended abstracts on Human factors in computing systems - CHI EA '06* (pp. 748–753). New York, New York, USA: ACM Press. http://doi.org/10.1145/1125451.1125601

Evangelista, B. (2016). Robots greet Westfield mall shoppers in San Francisco, San Jose. Retrieved from https://www.sfgate.com/business/article/Robots-greet-Westfield-mall-shoppers-in-San-10631291.php

Felberbaum, Y., & Lanir, J. (2018). Better Understanding of Foot Gestures: An Elicitation Study. In *Proceedings of CHI 2018* (Vol. 36, pp. 218–224). Montreal, Quebec, Canada: ACM Press. http://doi.org/10.1145/3173574.3173908

Findlater, L., Lee, B. Q., & Wobbrock, J. O. (2012). Beyond QWERTY: Augmenting Touch-Screen Keyboards with Multi-Touch Gestures for Non-Alphanumeric Input. In *Proceedings of CHI 2012* (pp. 2679–2682). Austin, Texas, USA: ACM Press.

Frisch, M., Heydekorn, J., & Dachselt, R. (2009). Investigating multi-touch and pen gestures for diagram editing on interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '09* (pp. 149–156). New York, New York, USA: ACM Press. http://doi.org/10.1145/1731903.1731933

Funes, M. M., Fortes, R. P. M., Trojahn, T. H., & Goularte, R. (2018). Gesture4All: a framework for 3D gestural interaction to improve accessibility of Web videos. In *In Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (pp. 2151–2158). Pau, France: ACM Press. http://doi.org/10.1145/3167132.3167363

Fusch, P. I., & Ness, L. R. (2015). Are We There Yet? Data Saturation in Qualitative Research. *The Qualitative Report*, *20*(9), 1408–1416.

Gerling, K. M., Livingston, I. J., Nacke, L. E., & Mandryk, R. L. (2012). Full-Body Motion-Based Game Interaction for Older Adults. In *CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1873–1882). Austin, Texas, USA: ACM Press.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin Co.

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory* (First). Routledge.

Goel, M., Findlater, L., & Wobbrock, J. O. (2012). WalkType: Using Accelerometer Data to Accommodate Situational Impairments in Mobile Touch Screen Text Entry. In *Proceedings of CHI 2012*. Austin, Texas, USA: ACM Press.

Gorilla Arm. (n.d.). Retrieved from http://www.catb.org/jargon/html/G/gorilla-arm.html

Grace, K., Grace, S., Maher, M. Lou, Mahzoon, M. J., Lee, L., LoCurto, L., & Outcault, B. (2017). The willful marionette: Exploring responses to embodied interaction. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition - C&C '17*. http://doi.org/10.1145/3059454.3059475

Grace, K., Wasinger, R., Ackad, C., Collins, A., Dawson, O., Gluga, R., … Tomitsch, M. (2013). Conveying Interactivity at an Interactive Public Information Display. In *PerDis '13* (pp. 329–332). Mountain View, CA: ACM Press. http://doi.org/10.1145/2541016.2541091

Grijincu, D., Nacenta, M., & Kristensson, P. O. (2014). User-defined Interface Gestures: Dataset and Analysis. In *2014 ACM Interactive Tabletops and Surfaces* (pp. 25–34). New York, NY, USA: ACM Press.

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82. http://doi.org/10.1177/1525822X05279903

Hachaj, T., & Ogiela, M. R. (2014). Rule-based approach to recognizing human body poses and gestures in real time. *Multimedia Systems*, *20*(81), 81–99. http://doi.org/10.1007/s00530-013-0332-2

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Hincapié-Ramos, J. D., Guo, X., & Irani, P. (2014a). The Consumed Endurance Workbench : A Tool to Assess Arm Fatigue during Mid-Air Interactions. In *Proceedings of DIS 2014* (pp. 109–112). Vancouver, BC, Canada: ACM Press.

Hincapié-Ramos, J. D., Guo, X., & Irani, P. (2014b). The consumed endurance workbench. In *Proceedings of the 2014 companion publication on Designing interactive systems - DIS Companion '14*. http://doi.org/10.1145/2598784.2602795

Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., & Irani, P. (2014). Consumed Endurance: A metric to quantify arm fatigue of mid-air interactions. In *SIGCHI Conference on Human Factors in Computing Systems*. http://doi.org/10.1145/2556288.2557130

Hincapié-Ramos, J., Guo, X., Moghadasian, P., & Irani, P. (2014). Consumed Endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of CHI 2014* (pp. 1063–1072). Toronto, ON, Canada: ACM Press. http://doi.org/10.1145/2556288.2557130

Hoff, L., Hornecker, E., Bertel, S., Weimar, B., & Media, F. (2016). Modifying Gesture Elicitation: Do Kinaesthetic Priming and Increased Production Reduce Legacy Bias? In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 86–91). New York, New York, USA: ACM Press. http://doi.org/dx.doi.org/10.1145/2839462.2839472

Hoggan, E., Nacenta, M. A., Kristensson, P. O., Williamson, J., Oulasvirta, A., & Lehtiö, A. (2013). Multi-Touch Pinch Gestures - Performance and Ergonomics. In *Proceedings of the International Conference on Interactive Tabletops and Surfaces (ITS'13).* http://doi.org/10.1145/2512349.2512817

Hoggan, E., Williamson, J., Oulasvirta, A., Nacenta, M., Kristensson, P. O., & Lehtiö, A. (2013). Multi-Touch Rotation Gestures: Performance and Ergonomics. In *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. http://doi.org/10.1145/2512349.2512817

Jacob, R. J. K. (1991). The Use of Eye Movements in Human-Computer interaction Techniques: What You Look At is What You Get. *ACM Transactions on Information Systems*, *9*(3), 152–169.

Jang, S., Stuerzlinger, W., Ambike, S., & Ramani, K. (2017). Modeling Cumulative Arm Fatigue in Mid-Air Interaction based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. http://doi.org/10.1145/3025453.3025523

Johnson-Glenberg, M. (2018). Immersive VR and Education: Embodied Design Principles that Include Gesture and Hand Controls. *Frontiers in Robotics and AI*, *5*(July), 81. http://doi.org/10.3389/FROBT.2018.00081

Jones, E., Alexander, J., Andreou, A., Irani, P., & Subramanian, S. (2010). GesText : Accelerometer-Based Gestural Text-Entry Systems. In *Proc. of CHI '10* (pp. 2173–2182). Atlanta, GA, USA: ACM Press.

Kane, S. K., Wobbrock, J. O., & Ladner, R. E. (2011). Usable Gestures for Blind People: Understanding Preference and Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 413–422). Vancouver, BC, Canada: ACM Press. http://doi.org/10.1145/1978942.1979001

Karam, M., & Schraefel, M. C. (2005). A Taxonomy of Gestures in Human Computer Interactions. http://doi.org/10.1.1.97.5474

Kendon, A. (1986). Current Issues in the Study of Gesture. In J.-L. Nespoulous, P. Perron, & A. Roch Lecours (Eds.), *The biological foundations of gestures: Motor and semiotic aspects* (pp. 23–47).

Kobayashi, M., Hiyama, A., & Miura, T. (2011). Elderly User Evaluation of Mobile Touchscreen Interactions. In *IFIP conference on human-computer interaction* (pp. 83–99).

Springer Berlin Heidelberg.

Kölsch, M., Beall, A. C., & Turk, M. (2003). The Postural Comfort Zone for Reaching Gestures. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. http://doi.org/10.1177/154193120304700427

Köpsel, A., & Bubalo, N. (2015). Benefiting from legacy bias. *Interactions*, *22*(5), 44–47. http://doi.org/10.1145/2803169

Krueger, M. W. (1993, July). Environmental Technology Making the Real Virtual. *Communications of the ACM*, *36*(7), 36–37.

Krupka, E., Karmon, K., Bloom, N., Freedman, D., Gurvich, I., Hurvitz, A., … Bar Hillel, A. (2017). Toward Realistic Hands Gesture Interface: Keeping it Simple for Developers and Machines. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)* (pp. 1887–1898). New York, NY, USA: ACM Press. http://doi.org/10.1145/3025453.3025508

Kunz, A., Schnelle-Walka, D., Alavi, A., Pölzer, S., Mühlhäuser, M., & Miesenberger, K. (2014). Making Tabletop Interaction Accessible for Blind Users. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces - ITS '14* (pp. 327–332). Dresden, Germany: ACM Press. http://doi.org/10.1145/2669485.2669541

Laban, R., & Ullmann, L. (1971). *The mastery of movement*.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University Of Chicago Press.

Lascarides, A., & Stone, M. (2009). A Formal Semantic Analysis of Gesture. *Journal of Semiotics*, *26*, 393–449. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.3741

Law, L. A. F., & Avin, K. G. (2010). Endurance time is joint-specific: A modelling and meta-analysis investigation. *Ergonomics*, *53*(1), 109–129. http://doi.org/10.1016/j.fusengdes.2017.02.047

Lenman, S., Bretzner, L., & Thuresson, B. (2002). Using marking menus to develop command sets for computer vision based hand gesture interfaces. *Proceedings of the Second Nordic Conference on Human-Computer Interaction - NordiCHI '02*. http://doi.org/10.1145/572043.572055

Liu, M., Nancel, M., & Vogel, D. (2015). Gunslinger : Subtle Arms-down Mid-air Interaction. In *Proceeedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. http://doi.org/10.1145/2807442.2807489

Liu, X., & Thomas, G. W. (2017). Gesture Interfaces: Minor Change in Effort, Major Impact on Appeal. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. http://doi.org/10.1145/3025453.3025513

Loke, L., Larssen, A. T., Robertson, T., & Edwards, J. (2007). Understanding movement for interaction design: frameworks and approaches. *Personal and Ubiquitous Computing*, *11*(8), 691–701. http://doi.org/10.1007/s00779-006-0132-1

Long, A. C., Landay, J. A., & Rowe, L. A. (1999). Implications for a gesture design tool. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, (January), 40–47. http://doi.org/10.1145/302979.302985

Maher, M. Lou, & Lee, L. (2017). *Designing for Gesture and Tangible Interaction*. (J. M. Carroll, Ed.) (Synthesis). Morgan & Claypool Publishers.

Majaranta, P., Aoki, H., Donegan, M., Hansen, D. W., & Hansen, J. P. (2011). *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies*. (P. Majaranta, Ed.). IGI Global.

Malu, M., Chundury, P., & Findlater, L. (2018). Exploring Accessible Smartwatch Interactions for People with Upper Body Motor Impairments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (pp. 1–12). Montreal, Quebec, Canada: ACM Press. http://doi.org/10.1145/3173574.3174062

Matsumoto, D., & Hwang, H. C. (2013). Cultural Similarities and Differences in Emblematic Gestures. *Journal of Nonverbal Behavior*, *37*(1), 1–27. http://doi.org/10.1007/s10919-012-0143-8

McAweeney, E., Zhang, H., & Nebeling, M. (2018). User-Driven Design Principles for Gesture Representations. In *Proceedings of CHI 2018* (p. 547:1-547:13). Montreal, Quebec, Canada: ACM Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press.

McNeill, D., & Levy, E. T. (1982). Conceptual representations in language activity and gesture. *Speech, Place, and Action. Studies in Deixis and Related Topics*, (January), 271–295.

Micire, M., Desai, M., Courtemanche, A., Tsui, K. M., & Yanco, H. a. (2009). Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '09* (pp. 41–48). New York, New York, USA: ACM Press. http://doi.org/10.1145/1731903.1731912

Moen, J. (2005). Towards people based movement interaction and kinaesthetic interaction experiences. In *Proceedings of the 4th decennial conference on Critical computing between sense and sensibility - CC '05* (p. 121). http://doi.org/10.1145/1094562.1094579

Morris, M. R. (2012). Web on the Wall: Insights from a Multimodal Interaction Elicitation

Study. In *ITS '12 Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (pp. 95–104). Cambridge, Massachusetts, USA: ACM Press.

Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., & Wobbrock, J. O. (2014). Reducing legacy bias in gesture elicitation studies. *Interactions*, *21*(3), 40–45.

Morris, M. R., Wobbrock, J. O., & Wilson, A. D. (2010). Understanding Users ' Preferences for Surface Gestures. In *Graphics Interface 2010* (pp. 261–268). Canadian Information Processing Society.

Morse, J. M. (1994). Designing funded qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of Qualitative Research* (Second Edi). Thousand Oaks, CA: SAGE Publications.

Mouloua, S. A., Mouloua, M., Mcconnell, D. S., & Hancock, P. A. (2018). The Effects of Handedness and Dominance on Motor Task Performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1237–1241).

Nacenta, M. A., Kamber, Y., Qiang, Y., & Kristensson, P. O. (2013). Memorability of Pre-designed & User-defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (p. 1099). New York, New York, USA: ACM Press. http://doi.org/10.1145/2470654.2466142

Nacher, V., Jaen, J., Navarro, E., Catala, A., & González, P. (2015). Multi-touch gestures for pre-kindergarten children. *International Journal of Human Computer Studies*, *73*, 37–51. http://doi.org/10.1016/j.ijhcs.2014.08.004

Nguyen, Q., & Kipp, M. (2014). Orientation Matters: Efficiency of Translation-rotation Multitouch Tasks. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2013–2016). http://doi.org/10.1145/2556288.2557399

Ni, T., Bowman, D. A., & North, C. (2011). AirStroke : Bringing Unistroke Text Entry to Freehand Gesture Interfaces. In *Proc. of CHI '11* (pp. 2473–2476). Vancouver, BC, Canada: ACM Press.

Nielsen. (2011). Nielsen Estimates Number of U.S. Television Homes to be 114.7 Million. Retrieved from http://www.nielsen.com/us/en/insights/news/2011/nielsen-estimates-number-of-u-s-television-homes-to-be-114-7-million.html

Nielsen, J. (1993). Iterative User-Interface Design. *IEEE Computer*.

Nielsen, J. (2004). Card Sorting: How Many Users to Test. Retrieved from https://www.nngroup.com/articles/card-sorting-how-many-users-to-test/

Nielsen, J. (2010). Why You Only Need to Test with 5 Users. Retrieved from https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Nielsen, J. (2012). How Many Test Users in a Usability Study? Retrieved from https://www.nngroup.com/articles/how-many-test-users/

Nielsen, J., & Landauer, T. K. (1993). A Mathematical Model of the Finding of Usability Problems. In *CHI '93 Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (pp. 206–213). Amsterdam, The Netherlands: ACM Press. http://doi.org/10.1145/169059.169166

Nielsen, M., Störring, M., Moeslund, T. B., & Granum, E. (2004). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction* (pp. 409–420). Springer Berlin Heidelberg.

Norman, D. A. (2010). Natural user interfaces are not natural. *Interactions*, *17*(3), 6. http://doi.org/10.1145/1744161.1744163

Norman, D. A. (2013). *The Design of Everyday Things*. Basic Books.

North, C., Dwyer, T., Lee, B., Fisher, D., & Isenberg, P. (2009). Understanding Multi-touch Manipulation for Surface Computing, 236–249.

O'Reilly, M., & Parker, N. (2012). 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research*, *13*(2), 190–197. http://doi.org/10.1177/1468794112446106

Oh, J. H., Jung, Y., Cho, Y., Hahm, C., Sin, H., & Lee, J. (2012). Hands-up: motion recognition using kinect and a ceiling to improve the convenience of human life. In *Proceedings of CHI 2012 Extended Abstracts* (pp. 1655–1660). ACM Press. http://doi.org/10.1145/2212776.2223688

Oh, U., & Findlater, L. (2013). The challenges and potential of end-user gesture customization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. New York, New York, USA: ACM Press. http://doi.org/10.1145/2470654.2466145

Pavlovic, V. I., Sharma, R., & Huang, T. S. (1997). Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *EEE Transactions on Pattern Analysis & Machine Intelligence*, *7*, 677–695.

Picard, R. W. (n.d.). What Happened to the Q Sensor? Retrieved from https://affect.media.mit.edu/projectpages/iCalm/iCalm-2-Q.html

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., … Ansari, R. (2002). Multimodal Human Discourse: Gesture and Speech. *ACM Transactions on Information SystemsComput.-Hum. Interact.*, *9*(3), 171–193. http://doi.org/10.1145/568513.568514

Rajanna, V., & Hammond, T. (2018). A Gaze Gesture-Based Paradigm for Situational Impairments, Accessibility, and Rich Interactions. In *Proceedings of the 2018 ACM*

Symposium on Eye Tracking Research & Applications (p. 102). Warsaw, Poland: ACM Press. http://doi.org/10.1145/3204493.3208344

Read, B. (2017). Rise of the airport robots. Retrieved from https://www.aerosociety.com/news/rise-of-the-airport-robots/

Robbe-Reiter, S., Carbonell, N., & Dauchy, P. (2000). Expression constraints in multimodal human-computer interaction. In *Proceedings of the 5th international conference on Intelligent user interfaces - IUI '00* (pp. 225–228). New York, New York, USA: ACM Press. http://doi.org/10.1145/325737.325852

Robbe, S. (1998). An Empirical Study of Speech and Gesture Interaction : Toward the Definition of Ergonomic Design Guidelines expression, (April), 349–350.

Ruiz, J., Li, Y., & Lank, E. (2011). User-defined motion gestures for mobile interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11* (pp. 197–206). New York, New York, USA: ACM Press. http://doi.org/10.1145/1978942.1978971

Ruiz, J., & Vogel, D. (2015). Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In *Proceedings of CHI 2015* (pp. 3347–3350). Seoul, Republic of Korea: ACM Press.

Rust, K., Malu, M., Anthony, L., & Findlater, L. K. (2014). Understanding Child-Defined Gestures and Children's Mental Models for Touchscreen Tabletop Interaction. In *IDC '14 Proceedings of the 2014 conference on Interaction design and children* (pp. 201–204). Aarhus, Denmark: ACM Press. http://doi.org/10.1145/2593968.2610452

Ryall, K., Morris, M. R., Everitt, K., Forlines, C., & Shen, C. (2006). Experiences with and Observations of Direct-Touch Tabletops. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)* (pp. 89–96). Adelaide, SA, Australia: IEEE Comput. Soc.

Schieber, F. (2003). Human factors and aging: Identifying and compensating for age-related deficits in sensory and cognitive function. In N. Charness & K. W. Schaie (Eds.), *Impact of technology on successful aging* (pp. 42–84). Springer Publishing Company.

Schuler, D., & Namioka, A. (1993). *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates Inc.

Sears, A., Lin, M., Jacko, J., & Xiao, Y. (2003). When computers fade: Pervasive computing and situationally-induced impairments and disabilities. *HCI International*, *2*(03), 1298–1302.

Sharp, T., Keskin, C., Robertson, D., Taylor, J., & Shotton, J. (2015). Accurate , Robust , and Flexible Real-time Hand Tracking. In *Proceedings of CHI 2015* (pp. 3633–3642). Seoul, Republic of Korea: ACM Press.

Siddhpuria, S., Katsuragawa, K., Wallace, J. R., & Lank, E. (2017). Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*. http://doi.org/10.1145/3064663.3064695

Sikora, C., & Burleson, W. (2017). The Dance of Emotion: Demonstrating Ubiquitous Understanding of Human Motion and Emotion in Support of Human Computer Interaction. In *Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, TX, USA: IEEE Comput. Soc.

Soegaard, M. (2019). Accessibility: Usability for all. Retrieved from https://www.interaction-design.org/literature/article/accessibility-usability-for-all

Spiro, I. (2012). Motion chain: a webcam game for crowdsourcing gesture collection. In *CHI'12 Extended Abstracts on Human Factors in Computing*. ACM Press. Retrieved from http://dl.acm.org/citation.cfm?id=2212452

Stößel, C., & Blessing, L. (2010). Mobile Device Interaction Gestures for Older Users. In *Proceedings: NordiCHI 2010* (pp. 793–796). Reykjavik; Iceland: ACM Press.

Stößel, C., Wandke, H., & Blessing, L. (2010). Gestural Interfaces for Elderly Users: Help or Hindrance? In *International Gesture Workshop* (pp. 269–280). Springer Berlin Heidelberg.

Takahashi, D. (2012). Xbox 360 Surpasses 66M Sold and Kinect Passes 18M Units. Retrieved from https://venturebeat.com/2012/01/09/xbox-360-surpassed-66m-sold-and-kinect-has-sold-18m-units/

Terry, M., & Mynatt, E. D. (2002). Recognizing creative needs in user interface design. In *Proceedings of the fourth conference on Creativity & cognition - C&C '02* (pp. 38–44). New York, New York, USA: ACM Press. http://doi.org/10.1145/581710.581718

Tullis, T., & Wood, L. (2004). How Many Users Are Enough for a Card-Sorting Study? In *Usability Professionals Association (UPA) 2004 Conference*. Minneapolis, MN.

Vatavu, R.-D. (2012). User-defined gestures for free-hand TV control. In *Proceedings of the 10th European conference on Interactive tv and video - EuroiTV '12* (pp. 45–48). New York, New York, USA: ACM Press. http://doi.org/10.1145/2325616.2325626

Vatavu, R.-D., & Wobbrock, J. O. (2015). Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems* (pp. 1325–1334). Seoul, Republic of Korea: ACM Press. http://doi.org/10.1145/2702123.2702223

Vatavu, R.-D., & Wobbrock, J. O. (2016). Between-Subjects Elicitation Studies: Formalization and Tool Support. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3390–3402). San Jose, CA, USA: ACM Press. http://doi.org/10.1145/2858036.2858228

Vercruyssen, M. (1997). Movement control and speed of behavior. In A. D. Fisk & W. A. Rogers (Eds.), *Handbook of human factors and the older adult* (pp. 55–86). San Diego, CA, US: Academic Press.

von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. *Proceedings of the 2004 Conference on Human Factors in Computing Systems - CHI '04*, 319–326. http://doi.org/10.1145/985692.985733

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). Retrieved from http://gerentedemediado.blogspot.com.es/2018/03/la-comodidad-de-la-primaria.html

Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. a. (2005). Maximizing the guessability of symbolic input. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05* (pp. 1869–1872). New York, New York, USA: ACM Press. http://doi.org/10.1145/1056808.1057043

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the 27th International conference on Human factors in computing systems - CHI 09* (p. 1083). New York, New York, USA: ACM Press. http://doi.org/10.1145/1518701.1518866

Wobbrock, J. O., Myers, B. A., & Kembel, J. A. (2003). EdgeWrite: A Stylus-Based Text Entry Method Designed for High Accuracy and Stability of Motion. In *UIST '03* (pp. 61–70). Vancouver, BC, Canada: ACM Press. http://doi.org/10.16309/j.cnki.issn.1007-1776.2003.03.004

Wobbrock, J. O., Rubinstein, J., Sawyer, M. W., & Duchowski, A. T. (2008). Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08* (pp. 11–18). Savannah, Georgia, USA: ACM Press. http://doi.org/10.1145/1344471.1344475

Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted Metadata for Image Search and Browsing. In *Proceedings of CHI 2003* (pp. 401–408). Ft. Lauderdale, Florida, USA: ACM Press. http://doi.org/10.1145/642611.642681

Yesilada, Y., Harper, S., Chen, T., & Trewin, S. (2010). Small-device users situationally impaired by input. *Computers in Human Behavior*, *26*(3), 427–435. http://doi.org/10.1016/j.chb.2009.12.001

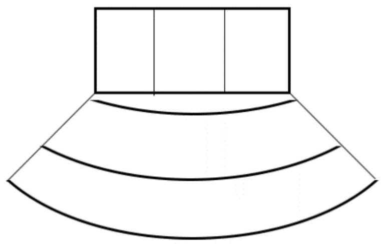
Zigelbaum, J., Browning, A., Leithinger, D., Bau, O., & Ishii, H. (2010). g-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface. In *Proceedings of TEI 2010* (pp. 261–264). Cambridge, Massachusetts, USA.

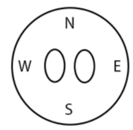
APPENDIX A

CODING FORM – STUDY 1

Participant No.:
Video name:
Gesture 1

Position





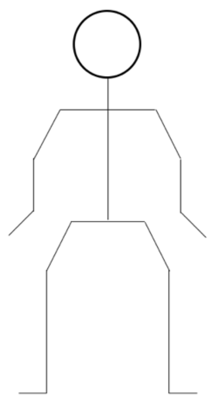
Common Primitives

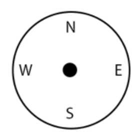
- ☐ Squat
- ☐ Jump
- ☐ Raise on toes

- ☐ Clap
- ☐ Snap
- ☐ Voice
- ☐ Gaze
- ☐ Grab/Clench
- ☐ Drag + Drop

Motion

- ☐ Swipe
- ☐ Punch
- ☐ Kick
- ☐ Point
- ☐ Circular
- ☐ Twist
- ☐ Step
- ☐ Lean/Tilt
- ☐ Turn
- ☐ Tap





Type of Gesture

- ☐ Absolute
- ☐ Relative

- ☐ Sequence
- ☐ Continuous
- ☐ Discrete
- ☐ Repeated

Speed

- ☐ Based on movement speed
- ☐ Based on position

Body Part

- ☐ Hand
- ☐ Forearm
- ☐ Elbow
- ☐ Arm
- ☐ Waist
- ☐ Foot
- ☐ Lower Leg
- ☐ Knee
- ☐ Leg
- ☐ Head
- ☐ Full body

Which side?

- ☐ Right
- ☐ Left

Person is:

- ☐ Standing
- ☐ Sitting

Hand Configuration

- ☐ 1 Finger
- ☐ 2 Finger
- ☐ Flat hand
- ☐ Fist

Palm position

- ☐ Up
- ☐ In
- ☐ Out
- ☐ Down
- ☐ Forward

Gesture Mapped to….

- ☐ Body ____
- ☐ Room ____
- ☐ Other ____

Notes:

APPENDIX B

DEMOGRAPHIC SURVEY – STUDY 2

Participant Number: ____________

Gesture Elicitation Study - Background Survey

1) Are you technologically savvy (in other words, do you program, build anything on the web, develop electrical devices, and/or solve problems with technology?)
    a. Yes
    b. No
    c. Considered myself savvy but not a computer, mechanical, or electrical engineer.

2) Do you own and/or currently use any tablet, video game systems (e.g. Wii, Xbox, etc.), smartphone, laptop, computer, and/or music player with touch screen?
    a. Yes
    b. No

3) Have you ever used a gesture recognition system before like Xbox Kinect and Leap Motion?
    a. Yes
    b. No

4) Do you currently experience any muscle or joint pain?
    a. Yes
    b. No
    c. Some

5) Do you like to work-out whenever you find the time and opportunity?
    a. Yes
    b. No
    c. Sure, mostly depends on my mood.

6) How long would you say that you have used a tablet or smartphone (not meant to be a trick question)?
    a. Since it came out
    b. About 1 to 2 years
    c. 3 and more
    d. Never

APPENDIX C

USER SURVEY – STUDY 2

Participant Number: ____________

Gesture Elicitation Study - User Survey

Question 1: For this task, please rank your preference for each gesture in an ascending order, with 1 being most preferred.

____

Question 2: What was the reason or factor for your ranking? Ranking was based on:

a. Comfort
b. Boost in energy or mood
c. Intuitiveness for this task (in other words, suitable and brought to mind first)
d. Creativity
e. Ease of communication with computational devices (easy for computers to recognize gestures)
f. Others: ____

Question 3: Imagine doing the same gestures repeatedly for an hour. Which of your gestures do you think would be most stressful to you?

Please list them here.

____

Question 4: What do you like about this task? Please select all that applies to you.

a. Easy to understand
b. Easy to perform gestures for
c. Similar to technologies you have used in the past
d. Others: List other reasons for your preference.

____

Question 5: What do you dislike about this task?

a. Unsuitable for gestures
b. Unrealistic
c. Stressful
d. Confusing
e. Others: Please list other reasons here.

____

Question 6: Did you feel physically stressed out doing this task?

a. Yes
b. No
c. Others: ____

APPENDIX D

USER SURVEY – STUDY 3

Participant Number: ____________

Gesture Elicitation Study - User Survey

1. What is your gender?
    a. Female
    b. Male
    c. Other

2. What is your age?

3. What is your race (circle all that apply)?
    a. White or Caucasian
    b. Hispanic or Latino
    c. Black or African American
    d. Native American or American Indian
    e. Asian
    f. Pacific Islander
    g. Other (please specify): ______________

4. What cultures are you most influenced by?

5. Are you right handed, left handed, or no preference?
    a. Right
    b. Left
    c. No preference

6. Current degree being pursued?
    a. Bachelors
    b. Masters
    c. PhD
    d. Other (please specify): ____________________

7. What is your major?

8. Do you play video games?
    a. Yes
    b. No

9. If you play video games, how often (Skip this question if you do not play video games)?
    a. Less than once a week
    b. Once a week
    c. Daily, but less than 1 hour a day
    d. Several hours a day (1-5 hours a day)

10. Have you ever used a Microsoft Kinect or similar product (e.g. Leap motion or any other gesture tracking hardware) for video games?
    a. Yes
    b. No

11. If you have, how often?

12. On a scale of 1 to 5, how comfortable are you using gesture tracking hardware (e.g. Microsoft Kinect, Leap Motion, Nintendo Wii)?

Not at all comfortable		Neutral		Very comfortable
1	2	3	4	5

13. Have you used these devices in other contexts besides video games (e.g. to interact with a public kiosk, in a museum, etc.)? If so, how have you used them?

APPENDIX E

GESTURE PRIMITIVE DEFINITIONS – STUDY 3

ASL: Using defined ASL hand or arm gestures.

Blink: Blinking one's eyes.

BMI: Any mention of "just thinking about it" and the action happening. Short for Brain-machine interface.

Clap: Clapping your hands together. Much greater force than pinch and implies that the two hands make contact, which may or may not be true for a pinch. Usually includes the hands making a noise on contact.

Draw: Any gesture that involves drawing. E.g. Draw an "X" or a circle or draw a check mark.

Expand: Used for any movement where multiple body parts move away from one another after starting out together or next to one another. Opposite of pinch.

Flick: Flicking with your arms / hands. Fast, short movement done with force. Like flicking an ant or lint away.

Gaze: Any gesture involving the user looking at a particular location or item.

Grab: Going from an open handed or partially open hand configuration to a fist like one is holding onto something. Similar to how one would grab a physical item.

Hover: Moving and <u>holding</u> a body part in position for several seconds. This may or may not be towards a particular object.

Jump: Jumping with the legs or full body upwards. Usually used as a complementary gesture to squat.

Kick: Kicking with your legs. Usually done quickly and with force.

Lean: Change in weight distribution rather than a rotation of a joint. E.g. shifting weight onto the balls of your feet or shifting your weight from one foot to the other.

Move: Should be used in instances in which moving the body part to a specific position in meaningful to the user, for example when "dragging" or moving objects to a particular location. In cases in which the user moves their body part in preparation for a gesture, we ignore this movement. This primitive is also used in cases in which there's a meaningful expansion of the arm or leg that wouldn't be captured if it wasn't split out on its own.

Nod: Up / down "yes" nod or looking up / down by rotating your neck. Or left / right "no" nod or looking left / right by rotating your neck.

Pen Click: Only performed by one participant that used a pen as part of his gestures. He would click the pen open / closed to indicate things such as selection, etc.

Pinch: Used for any movement where multiple body parts move towards one another after being separated (e.g. bringing your arms together, or pinching two fingers together). Opposite of expand.

Point: Always to a specific location in space (i.e. absolute….whether it's an object, the edge of the screen, etc.) and doesn't have the force of a tap or necessarily the quick joint rotation. Shorter hold than a hover. This could include the participant specifying that they'd like to touch the screen.

Pull: Usually a slower movement with a lot of perceived force. Like you're pulling something towards you.

Push: Usually a slower movement with a lot of perceived force. Like you're pushing something away from you.

Rotate: Any joint rotation that doesn't fit into another, more meaningful, gesture primitive.

Scoop: Like picking something up with an arcing motion.

Shake: Short quick movements back and forth of a joint. Often times, this is called an erasing" gesture.

Shoot gun: Hand gesture like one's shooting a gun where the index finger is straight and pointed at something and the thumb bends to indicate the shooting action.

Slide: Moving a body part (usually arm or leg) from one location to another on a straight or arched path. More like a dragging motion instead of a swipe. This is slower than a swipe and has less speed / force associated with it.

Snap: Snapping one's fingers together. Short, quick movement, with sound.

Squat: Squatting or crouching. Usually involves only movement of the legs.

Step: One foot movement and landing, but not continuing with a second step.

Swipe: Moving a body part (usually arm or leg) from one location to another on a straight or arched path. Similar to swiping on a touch screen. This is faster than a slide and implies more speed / force.

Tap: Like a point but with a quick rotation of a joint and added force. This could include the participant specifying that they'd like to touch the screen.

Throw: Like physically throwing an object in a particular direction. Indicates that the user is "holding" something and that they're tossing it with force in a particular direction, usually in an arc.

Thumbs up: Thumbs up gesture / symbol. Usually with a hover / hold at the end.

Tilt: Tilting head with the ear going towards the shoulder and the face remaining facing forward.

Unsquat: Standing up straight from a squatting position.

Voice: Any sort of voice command.

Walk: Multiple steps in a row in which the user is targeting a specific spot.

X: Crossing two limbs in an X shape