

**ARE GESTURE INTERFACES OUT OF TOUCH? CHALLENGES WITH THE
ADOPTION OF EMERGING TECHNOLOGY**

by

Amanda Powell

B.A. Recreation and Business, University of Waterloo, 2007

A Major Research Paper
presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Master of Digital Media
in the program of Digital Media

Toronto, Ontario, Canada, 2017

© Amanda Powell, 2017

Author's Declaration

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

ARE GESTURE INTERFACES OUT OF TOUCH? CHALLENGES WITH THE
ADOPTION OF EMERGING TECHNOLOGY

Master of Digital Media, 2017

Amanda Powell

Master of Digital Media, Ryerson University

Abstract

Many studies in the field of human-computer interaction (HCI) point to gesture-based interaction (GBI) as a transformative method for communicating with computers. GBI allows people to use common body language like waving or pointing to manipulate devices without physically touching them. Current research suggests that moving beyond traditional mechanical devices such as a mouse or keyboard may create richer and more ‘natural’ user experiences. Despite this finding, this mode of interaction has not seen broadscale adoption. A critical analysis of the work taking place in both industry and HCI studies demonstrates tension between the theory and practice of creating gesture-centric interfaces. This study provides a critical overview of the issues, technological, design and social, that pose a challenge to the widescale adoption of this technology.

Acknowledgments

This major research paper would not have been possible without the guidance and thoughtful advice from my supervisor, Dr. Jason Boyd. I would also like to thank Dr. Ali Mazalek, who provided valuable feedback as a second reader.

I would like to acknowledge the faculty and staff of the MDM program for facilitating a stimulating, supportive and rewarding academic experience. Moreover, I would like to express my sincere gratitude to my fellow MDM candidates who have believed in me, inspired me and made my time at Ryerson University worthwhile.

Dedication

In dedication to my partner, Jon, whose love (and cooking) fuels my endeavours.

Table of Contents

Author's Declaration	ii
Abstract	iii
Acknowledgments	iv
Dedication	v
Table of Contents	vi
Chapter 1: Introduction	1
Chapter 2: Background	5
What is Human-Computer Interaction?	5
The User Interface Revolution	6
Gesture-based Interaction	8
Chapter 3: Applications	10
Gaming and Entertainment Application	10
Applications in Healthcare	14
Assistive Technology	18
Chapter 4: Is Theory at Odds with Practice?	22
Embodied Interaction	22
Natural Interaction	23
Theory Versus Practice	28
Chapter 5: A 'NUI' Set of Issues	30
Technological Challenges	30
Design Challenges	32

Social Challenges	35
Chapter 6: Conclusion	39
References	41

Introduction

One of the promises of digital technology is to improve our lives. People have imagined—and have come to expect—a digital future where technology seamlessly integrates into our environment and can facilitate the undertaking of the most mundane tasks with effortless direction. While the progression of technological innovation has been dizzying, we are still subject to ineffectively designed systems that often add to, rather than remove, barriers to completing tasks in an optimal manner.

Not that long ago, computers could only complete tasks through detailed and complex instructions from people. The technology that existed at the dawn of the age of computing was not advanced enough to understand us, so we had to change our way of thinking in order to understand the technology. Computer engineers created an alternative ‘language’ of interactions that bridged the gap between human and computer. People were forced to read detailed manuals, memorize precise commands according to the computer's parameters and undergo significant training to understand how to interact with a computer.

As technological innovations evolve, so too has our relationship with technology. We can be more ‘ourselves’ and the computer can more readily understand our needs. Technology is smart enough—or becoming smart enough—to reverse the relationship humans originally had with computers and we can increasingly rely on their computing power to solve our problems. Currently, new technologies are built with the human (and human needs) at the heart of their designs, making them increasingly intuitive to interact with.

The field of human-computer interaction (HCI) studies user interface design and explores novel interaction methods that can support the creation of useful and usable systems for people. Consider for a moment just a few of the activities that you do each day with the help of a

computer. You might check your email and social media accounts, listen to a podcast, draw, or monitor your heartrate while you exercise. The list is nearly infinite. Now consider how you navigated each device or system. You may have used a mouse and keyboard to read your email, touched and tapped a screen to check social media on your smartphone, used a stylus or pen device to draw on a tablet, and relied on body-based sensors to monitor your heartrate while you worked out. Indeed, much of our interaction with computers has become so commonplace, we rarely think of the complexities they involve. While these interactions became natural over time, each involved a learning curve when they were new. What if we could develop computing systems that required no such learning curve? Researchers in the field of HCI suggest that moving beyond all of the above interaction modes present opportunities to utilize human communication skills like speech, gestures and body language to create more natural and intuitive experiences for computer users. Natural user interfaces (NUIs) are hardware and software systems designed to make the interaction between human and computer feel as intuitive and seamless as possible. Gestures, such as waving and pointing, are one possible input modality of natural user interfaces. Norman (2010) identifies gesture as a potential method to improve interactions where traditional inputs like a keyboard and mouse are inconvenient.

The topic of this paper was inspired by my own inconvenience of using existing computing technology in the kitchen. Like many others, I find inspiration and recipes online and access this information through a smartphone or tablet. Needless to say, the kitchen is hardly an ideal environment for these devices. Messy hands limit my use of touchscreen devices while cooking, requiring me to stop and wash my hands to read the next step in the recipe. I refused to believe that I was the only person to have experienced this challenge. Surely there had to be a hands-free tool that could help. The answer to my problem is gesture recognition systems which

can track a user's body motions and translate them into instructions for the computer.

Technology that utilizes gesture recognition currently exists in commercially available products such as Microsoft's Kinect, the Leap Motion Controller (LMC) and the Myo Gesture Control Armband. Furthermore, the successful application of these technologies can be seen in niche areas like healthcare, gaming and assistive technology for people with disabilities. However, no one has applied this technology to the household kitchen or other spaces in which people commonly live and work.

This presents a perplexing situation. Existing research on gesture-based interaction is premised on the idea that computer interaction that mirrors human-to-human communication can make using new systems easier. Yet this research has mainly focused on building the necessary technology, and, perhaps as a result, has often failed to give enough consideration to why it is useful, how it will be used (Rico & Brewster, 2010) or what new challenges it can bring with it (Norman, 2010). Applications in industry demonstrate that gesture-based interaction has been effective in solving problems that traditional interactive tools cannot. Given this success, the question becomes why, especially given that gesture-based technologies currently exist in the consumer market and are affordable for many, there is no widescale adoption in everyday life? As a result of this question, this paper focuses on the *human* aspect of this technology—the motivations, experiences and barriers that frame how people think about and use computers. While there have been successful adoptions of this technology in specialized fields/environments (such as gaming, healthcare and assistive technology), the applications outside of these specific areas are presently limited. My research addresses this perplexing situation by outlining a set of challenges that currently limit the nature and extent of the adoption of gesture-based interaction in everyday activities. In no way is this work intended to be predictive in nature, rather it is an

attempt to identify, through the examination of previous successes and failures, the inevitable issues that are inherent with this technology. There are three areas that are presenting barriers to the adoption of this technology outside of specialized fields/environments: technological, design and social.

Before examining these barriers in more detail, this paper will provide a summary of developments from existing HCI work to further understand the problem with existing interaction methods and the benefits of using gesture-based interaction. Next, a critical analysis of the work taking place in both industry and HCI studies demonstrates a disconnect between the theory and practice of creating gesture systems. Finally, as a result of examining industry applications, I identify the unique technological, design and social challenges that need to be resolved before widespread adoption of this technology can take place, and before I will be able to solve my own culinary conundrum with using technology in the kitchen.

Background

What is Human-Computer Interaction?

Until the late 1970s, computing was largely confined to small groups of skilled technologists. It was not until the emergence of ‘personal computing’—the growth of the non-technical user—that computer scientists began to evaluate their work using concepts like ‘usability’ and ‘efficiency’. The field of human-computer interaction (HCI) was established in the 1980s when computer engineers began to consider the more ‘human’ areas of computing, such as cognitive science and human factors engineering, in an effort to improve computer interaction (Carroll, n.d.). The key goal of the work of HCI researchers is to observe humans using computers and, extrapolating from that observation, develop design principles for the creation of usable systems. Most commonly, HCI outcomes are centred on creating the most effective, efficient, usable or useful way to navigate computers, but many also consider factors like pleasure, fun and even ‘naturalness’.

User Experience (UX) extends the concept of HCI to include not only how to design systems that function well, but also considers how a user feels when using the system or product. The user’s feelings and motivations are considered throughout the product lifecycle and might include, for example, the process of buying a product, opening the packaging, learning to use it for the first time and even telling others about it. Companies have been turning to user experience to gain competitive advantage, with Apple being one of the more notable companies to embrace this design perspective.

Until the mid 1990s, HCI was focused on improving the personal computer experience; however, technological developments over the past decade have encouraged an expansion of the field and the introduction of novel approaches to computer interaction beyond that developed for

the personal computer, including gesture-based, voice-based and even brain-computer interaction (Preece, Rogers, & Sharp, 2015).

The User Interface Revolution

A user interface (UI) is considered to be the activity space where the interaction between human and computer occurs. Interfaces can be classified a number of ways: by their function (to be invisible or predictive, or ‘smart’), by their interaction method (command line, graphical), by their input mode (gesture, voice) or by the type of platform they are designed for (mobile, wearable) (Preece et al., 2015). How to define a UI is often confusing because these classifications are not mutually exclusive, nor are their definitions clear (as will be seen in the discussion of natural user interfaces in chapter four). Several studies provide a comprehensive review of the evolution of HCI and UI (Myers, 1998; Preece et al., 2015); however, this paper will focus on three important interface developments that contributed to the development of gesture-based interaction: command-line, graphical user and natural user interfaces.

The command-line interface (CLI) is one of the earliest versions of a UI, requiring users to type commands that direct the computer to complete a task. One of the challenges of CLI is that it requires users to commit the commands to memory or refer to a manual for a list of options. Systems that use CLIs were not intuitive, requiring specialty knowledge to operate, and therefore they failed to earn widespread adoption.

The introduction of direct manipulation and the mouse in the 1960s (Myers, 1998) led to more useful and easily discoverable interactions where the user could point a mouse at an object (such as an icon) on a graphical user interface (GUI) to select, open and move items without needing to memorize a series of text-based commands. The graphical user interface, sometimes

used interchangeably with WIMP (windows, icons, menus, pointer) relied heavily on direct (mouse-based) manipulation of visuals, and, as a result, reduced barriers to computing for non-technical users. Some of the most common technology available to us today, the smartphone, mostly employs a GUI style, where the interaction input swaps a mouse with the touch of a finger. While it is true that the popularity of GUIs in personal computing has largely supplanted CLIs for the average computer user, typed commands can still be seen in specialized applications such as computer programming languages and Google's search bar.

Mark Weiser (1991), previously the chief scientist at Xerox PARC (Palo Alto Research Center), suggested, in his seminal article *The Computer for the 21st Century*, that the era of desktop computing would be a transition towards achieving the real potential of computing. Further, he claims that he and his fellow colleagues at PARC were “trying to conceive a new way of thinking about computers, one that takes into account the human world and allows the computers themselves to vanish into the background” (Weiser, 1991, p. 94). The natural user interface (NUI) is one modern embodiment of the era that Weiser and colleagues were imagining while they researched the tools and practices that much of HCI was built upon.

The NUI is one that draws on ‘natural’ human communication skills like gesture, touch and speech. Researchers suggest that using familiar communication tools, rather than learning to use a keyboard or mouse, can result in the interface being more intuitive and easier to learn (Preece et al., 2015). NUIs represent a model that can potentially further reduce the barriers to computing and can contribute to making everyday computing tasks more achievable.

The introduction of the GUI did not eliminate the CLI, it just allowed the text-based interface to find its best application. Similarly, claims that the NUI will replace the GUI are likely too ambitious. The GUI will remain where its affordances are best applied, most likely in

desktop computing, and give space to allow the NUI to improve systems where the GUI is not the right fit. This revolution of interface design demonstrates the value of having a variety of ways to interact with computers. Not all tasks and contexts require the same tools and by allowing each interface to find their niche application, we benefit from these distinct approaches.

Gesture-based Interaction

Many studies in the field of HCI point to gesture-based interaction (GBI) as a transformative mode by which users can perform expressive body movements to communicate with computers. The term ‘gesture’ here simply means to use body movement to convey action or meaning. From this definition, both touch gestures like swiping and pinching and touchless gestures like pointing and waving fit this description. Touch gestures have been widely adopted through the pervasive use of touchscreen technology. As a result, this paper understands gesture-based interaction as including touchless gestures of the face, head, hands, arms and full body. Instead of using a mouse to manipulate an object, the user might rotate their hand from side to side to indicate a command. The concepts of GBI and NUIs have been paired together so often in research and application that it is hard to say ‘gesture-based’ without automatically assuming it is realized in a NUI. While one doesn’t define the other, GBI researchers make similar claims of the benefits of gesture to be in line with the ‘natural experience’ argument. GBI can be helpful in contexts when the user's hands are dirty or occupied, when accessibility and mobility present barriers in using a device, when the user prefers to interact with a device from across the room or where performative actions enhance social experiences.

Gestures have no significance to the computer unless the system can recognize the movement and translate its meaning. The gesture recognition process begins when the user

inputs a gesture, the sensor detects the gesture and deduces its meaning, and finally, the device performs the task initialized by the gesture and returns an output to signal completion (Bansal, 2016). Bansal (2016) further suggests the sensor type used to detect the user's meaning can be categorized into three broad categories: device (glove, pen), vision (camera, image capture) and electrical (measuring the distance between the object and body part). Earlier systems often used device-based technologies (i.e. adding sensors to a pair of gloves), but with developments in the field, vision-based tools that enable in-air gestures are becoming more common (van den Hoven & Mazalek, 2011)¹.

¹ A more detailed categorization of available gesture recognition tools and their uses can be found in van den Hoven and Mazalek (2011).

Applications

Some of the first applications to use gesture recognition were developed in the 1960s. Myers' (1998) HCI review points to an early system called AMBIT/G which used a combination of menu items, pointing devices, and gesture recognition among other interface elements. Myron Krueger's experimental VIDEOPLACE system in the 1970s created an augmented reality space where live video reacted to users' body gestures (van den Hoven & Mazalek, 2011). In 2002, Steven Spielberg's *Minority Report* popularized a futuristic vision of human-computer interactions where in-air hand gestures are used to interact with a digital world around the user. However, it was not until late in the first decade of the 2000s when the first camera-based recognition system could be found in the homes of consumers around the world.

Gaming and Entertainment Application

The gaming and entertainment industries have sparked the development of commercially available gesture recognition hardware. Nintendo's Wii console, with Wiimote controller, was the first to launch in 2008 and cause a ripple effect through the gaming industry. Two years later, Sony and Microsoft, two of the largest players in the industry, launched gesture recognition devices for their respective consoles the PlayStation Move and Xbox 360. The release of these three systems sparked a craze in gaming where body movement that mimicked real world actions were used to control and interact with games. In living rooms around the world, players of all ages were swinging imaginary golf clubs and aiming their bowling ball down a digital lane in a virtual bowling alley. While all three consoles contributed to the popularity and adoption of gesture interaction in gaming, this paper will focus solely on the Kinect as it is the only device that does not rely on additional user input from hand-held controllers.

The first generation Kinect was launched in November 2010 with the tagline “You Are The Controller” (Weinberger, 2015). Not long after the launch, media outlets claimed the Kinect had earned the Guinness World Record’s title of “fastest-selling consumer electronic device” selling eight million devices in the first 60 days of its release (Kessler, 2011). The Kinect was designed for use with the Xbox console as an alternative to traditional controllers. While most consumers rushed home with their new toy to experience a revolutionary gaming experience, others were more eager to see how the technology could be applied to other applications by reverse engineering the system (Giles, 2010). The hacker and hobbyist community was buzzing when DIY electronic retailer, Adafruit, posted a \$1000 USD (later increased to \$3000 USD) reward for the first person who could share an open source driver that would connect the Kinect to any operating system (“The open Kinect project,” 2010). Not surprisingly, Microsoft’s early reactions to the bounty did not condone tampering with the system. The sour response from Microsoft only seemed to have fueled the hacker community’s motivation to crack the code, which they did in a matter of a few days. From robotic toilet flushing controls to music composition by stepping on invisible piano keys (Peckham, 2011), early inventions that utilized the Kinect’s gesture recognition system may not have been the most useful, but definitely earned points for creativity. It seems that researchers and hobbyists had long been waiting for a low-cost gesture recognition system to hit the market. After finally recognizing the value of contributions from hobbyists and researchers alike, Microsoft capitalized on the thirst for diversifying the use of the Kinect throughout the DIY community. In early 2011, the company announced the release of a software developer kit (SDK) with authorized code that opened the Kinect for non-commercial adaptations.

A unique Kinect application that has garnered much attention from the hobbyist community and Microsoft itself was an interactive puppet created by designer/artists Theo Watson and Emily Gobeille for their company Design I/O. Watson and Gobeille's creative use of the Kinect continued to develop over many years, and now they are renowned for their unique approach to creating interactive installations for children. One notable installation is a large-scale immersive, interactive ecosystem called Connected Worlds. Connected Worlds links multiple Kinects together so children can freely move throughout a 3000 square foot space enclosed by screen-based walls ("Connected worlds," n.d.). Children are encouraged to use body movements to plant seeds and watch them grow on the walls, and divert flowing 'water' to the areas within the installation that require hydration ("Connected worlds," n.d.). The interactive spaces created by Design I/O offers children places to learn and grow by leveraging both physical space and digital tools to facilitate an understanding of how their individual actions can impact a larger environment.

While initial sales of the Kinect spiked, by 2013 Microsoft was looking for ways to ensure that the Kinect continued to get into the hands of more players. Over the next two years, Microsoft coupled the Xbox 360 with the Kinect, forcing every new buyer to own both, but then decoupled the combination, offering the Xbox separately as a console with the Kinect offered as an add-on. At the same time, the game developer community could not keep up with the demand to create games that leveraged the interactive possibilities of the Kinect. Game developers were challenged by the fact that they needed to develop a product that could work with multiple input possibilities (i.e. a traditional controller and the Kinect). The Kinect offered gamers a whole new way of playing; however, it forced developers to rethink how their audiences would interact with their products (Hruska, 2014). In the end, most developers believed that this was far too costly in

time and effort, considering the effectiveness of the conventional controller (something that was not only familiar to them but to their audience as well). With a lack of interesting games to play, interest continued to dwindle from the gaming community for the device. It was clear that Microsoft was giving up on the device when the company removed the Kinect port from their Xbox console. A free adapter was later offered to legacy users who still wished to use their Kinect.

The surge of GBI home gaming systems within households across the globe demonstrated the enthusiasm and excitement by the gaming community to experiment with new styles of interaction that promised to immerse them further into ever-increasingly elaborate digital playgrounds. However, the lack of games that leveraged the Kinect were limited, which eventually led to the demise of the system. Microsoft had naive expectations that game developers had the interest and resources to invest in this new technology. Perhaps allocating their own in-house resources to game development would have demonstrated to other developers that the Kinect was worth investing in. Microsoft was the first major player to put this innovative technology in the hands of consumers, but they failed to provide useful ways for people to take advantage it. The Kinect may not have stood the test of time, but it should still be considered a major milestone in the adoption of GBI in gaming systems, even if its success was relatively short-lived.

In my opinion, the initial successful adoption of the Kinect was a result of two factors. First, using common, performative body movements creates a shared experience of the game for players and spectators. Non-players in the room become spectators and are often more interested in watching the player perform their actions than passively following what is happening on a screen. Perhaps being ‘next to the action’ made non-players feel more involved and invested in

seeing the player accomplish the goals of the game. This shared experience between player and spectator is more prominent in games where gestures are performative and expressive and not limited by the input of a handheld controller. This aspect of the interactions opened up gaming to new players who would not typically be considered 'serious gamers'.

The second factor that made the introduction of consumer GBI devices a success, was that learning took place through games. The system required players to learn a new way of communicating with their console that was largely unfamiliar to most people at the time. Masked by the need to learn mechanics and rules for any new game, the Kinect was able to reduce the barrier to learning the system.

The adoption of GBI technology in the gaming and entertainment industries can be linked to performative gestures that connected the performer and spectator through shared social experiences. The introduction of this interaction style through games and installations reduced barriers to learning and increased the immersion of both the performer and spectator. A lack of interesting games eventually led to the abandonment of the Kinect by the gaming community. However, the adoption of the Kinect by researchers and hobbyists presented opportunities for application outside of home gaming.

Applications in Healthcare

Healthcare practitioners, such as doctors and dentists, frequently rely on medical imaging and 3D models of their patients to guide them during surgical procedures. This information can be viewed with the assistance of joysticks, touchscreens, control panels or other tools that require the practitioner to physically touch a tool, however, this presents a challenge of maintaining sterility. Tools that require a practitioner's touch to operate can impede workflows and decrease

efficiency. Practitioners need to step away from the operating table to scrub out, use the tools and then scrub back in, wasting valuable operating room time—a cost of \$62 per *minute* (“GestSure,” n.d.). Alternatively, practitioners can rely on the support of their non-sterilized assistants, but verbally communicating the request can increase error rates, adding to further time wasted and increased frustration (Mewes, Hensen, Wacker, & Hansen, 2017).

Maintaining sterility is a unique challenge that can be addressed with gesture-based interaction. Several feasibility studies have been conducted to compare the efficiency and usability of GBI with conventional techniques in healthcare settings. Wipfli, Dubois-Ferrière, Budry, Hoffmeyer, and Lovis (2016) conducted an evaluation of the efficiency and ease of use of viewing medical images with GBI in a simulated surgical environment. The study compared three interaction methods: GBI with a Kinect, verbal task delegation to an assistant and direct manipulation using a mouse without a sterile cover. Efficiency was evaluated based on the elapsed time of each tested scenario. The results demonstrated that using a mouse was fastest, followed by GBI and verbal task delegation. Participants were given the opportunity to evaluate ease of use and perceived efficiency and this showed the results in the same order. Participants noted the use of gestures were more satisfying and resulted in fewer errors than when they provided verbal instructions to an assistant. Since using a non-sterile mouse in the operating room is impractical, GBI therefore offers a more efficient and satisfying experience compared to relying on the support of an assistant.

Hettig et al. (2017) presented contrasting results. They designed a comparative study to test two gesture recognition devices (LMC and Myo) with two existing techniques (task delegation and joystick) in a clinical sterile environment. The author chose to test the Leap Motion Controller (LMC) because of how small the device is and its ability to detect finger

motions and the Myo Gesture Control Armband (Myo) because it is a wearable device that can reduce location limitations. Ten experienced radiologists simulated a diagnostic neuroradiological treatment for two frequently used tasks. The unique part about this research is the comparison between two types of tasks. In the study, Hettig et al. (2017) compared the experience of radiologists accessing 2D images and a more complex task—3D volume renderings. The study evaluated the success of the interaction tools through three metrics: task completion time, perceived task difficulty and subjective workload. Overall, the existing operating room techniques (task delegation and joystick) performed the best to execute both tasks. Gesture inputs were rated as fast, sterile and intuitive to users, however they failed to exceed the established method. The Myo did show potential for the simplified 2D task, demonstrating the relevance of a task to the overall performance of each input method. The authors note that practitioners have more familiarity with existing techniques and that additional training or practice with the tools could have had better outcomes (Hettig et al., 2017).

In both experiments, the outcome demonstrated that existing techniques are more efficient than new GBI techniques. This presents an interesting dilemma. The existing techniques are not appropriate and cause lost money and disrupted workflow in the operating room, yet practitioners are not reporting that the new tools are efficient enough. Researchers profess that GBI is natural and intuitive, but when it does not meet these design goals, the results fall flat. Even with mixed results in the lab, gesture recognition systems are being created for commercial use in the operating room. GBI systems are not necessarily creating intuitive interactions for healthcare practitioners but they are solving one very specific issue: how to maintain sterility and allow practitioners to access digital patient information.

A rise in this research area can be seen to correspond with the market introduction of low-cost depth sensors. Mewes et al. (2017) note the increase of publications focusing on natural user interfaces in the medical domain increased from 2.2 published papers per year before 2013 to 10.3 papers published in an average year after the release of the Kinect SDK in 2012 and the LMC in 2013. The literature demonstrates a range of technical approaches, with Kinect and LMC being among the most common (Mewes et al., 2017). Prior to their release, researchers were building their own systems. Examples of this can be seen through the use of RGB colour cameras (Wachs et al., 2008), time of flight cameras (Soutschek, Penne, Hornegger, & Kornhuber, 2008) and body worn inertial sensors (Schwarz, Bigdelou, & Navab, 2011).

Therapixel, the creators of a touchless navigation system for medical use, claim that their proprietary system is 10x faster than asking an assistant to complete the task and 20x faster than scrubbing out to complete the task on their own (“Therapixel,” n.d.). These results significantly conflict with results in lab settings, however this could be a result of a better designed system or of practitioners having more time to learn the system before testing occurred. Tedcas, nz technologies, GestSure and Scopis have all created similar gesture-based systems to allow doctors to maintain sterility in the operating room while interacting with digital tools (“GestSure,” n.d., “nz technologies inc.,” n.d., “Scopis,” n.d., “TedCas,” n.d.). Both Tedcas and GestSure have made commercial use of the Kinect in their systems. Tedcas has also incorporated the LMC and Myo in their TedCube device that can create a bridge for any piece of healthcare equipment to be used through gesture control (“TedCas,” n.d.).

The arrival of numerous consumer products demonstrates the interest of the healthcare industry in adopting GBI. Feasibility studies conducted in the lab indicate that GBI can provide a solution for doctors to maintain sterility in the operating room. However, the interaction mode

impedes efficiency compared to traditional methods. Despite the increase in available products, Mewes et al. (2017) suggest there should be less of a focus on moving forward with technical developments and an increased effort in researching improvements and evaluating the usability of this technology in the medical field. It is through years of training and practice that doctors have established their techniques in the operating room. As a result, it is naive for researchers to assume that an ‘intuitive’ interface can replace years of existing procedure. In order to improve efficiency in the operating room, more attention needs to be given to ensure doctors are well trained and comfortable with new ways of interacting.

Assistive Technology

The application of gesture recognition technology can also be seen in HCI research and in the development of commercial products for assistive devices that support people with auditory, visual and physical impairments.

With its ability to accurately detect in-air hand gestures, researchers have used gesture recognition as a tool to convert sign language into speech and text. Similar to the medical community, researchers have been developing novel systems to recognize sign language for some time, using sensors and vision-based systems. The introduction of consumer products like the Kinect and LMC created new opportunities to advance the digital translation of American Sign Language (ASL) through computer recognition. Both devices have been used in accuracy tests aiming to build out the gesture recognition library for ASL (Mapari & Kharat, 2016; Rosa & Elizondo, 2014). Using devices to assist individuals with hearing impairments to communicate with others without the use of a sign language interpreter can help to break down the barriers these individuals face daily.

A proprietary hardware and software solution developed by MotionSavvy can translate ASL into grammatically correct speech using gesture recognition (“MotionSavvy,” n.d.). Their hardware tablet includes a front-facing camera that can recognize both hands for quick translation. The tablet, called UNI, is currently being piloted at the Rochester Airport in New York, USA and will be available to individuals once the pilot project testing has been completed. MotionSavvy’s website also suggests applications in retail, banking, medical and transportation, to help organizations communicate more effectively with their hearing impaired customers. The custom tablet can be purchased preloaded with their gesture recognition software, or buyers can opt for a more cost effective option to purchase the software for home use with the LMC. The key advantage of the tablet is its ability to take the technology anywhere it is needed, whereas combining the software with the LMC will restrict usage to a desktop or tabletop surface. Ideally, it would be beneficial if front-facing cameras in standard smartphones and tablets had the capacity to recognize gestures to the same extent. Users could then purchase the software for use on any mobile device, reducing the number of devices an individual needs to carry. Furthermore, this could help users with hearing impairments overcome their fears of being singled out as their need to use assistive devices would be greatly (or even completely) reduced. While this is impressive, the impacts of this technology can also be seen by the innovations being developed for those with physical disabilities.

Another lab-based Kinect adaptation has proposed a solution to the challenge of controlling a wheelchair by people who are unable to utilize their upper limbs (Kondori, Yousefi, Liu, & Li, 2014). Most electric wheelchairs available on the market use a joystick to directionally navigate the movement of the chair. Kondori et al. (2014) placed a Kinect at the front lower base of the wheelchair at an angle that allows the camera to view the user's head

movements. The software then maps the head gestures to the accompanying directional instruction to maneuver the wheelchair. The solution presents a series of challenges, such as having to carry around a Kinect all day and a safety concern that the Kinect will recognize gestures from people behind the wheelchair and misinterpret them as intended actions from the user. These challenges have been addressed by Sesame, the creator of the world's first touch-free phone designed for people with physical disabilities.

The Sesame phone is a modified Google Android smartphone that uses a front-facing camera to detect head gestures, allowing users to read email, browse the web and many other common smartphone tasks ("Sesame," n.d.). The device employs a voice command "open sesame" so the device knows when to activate and begin reading head gestures from the user. Head movements in combination with Sesame's computer vision algorithm allows users to modify all interaction features traditionally available with the touch of a finger like swiping, scrolling and pinching.

Despite the development of touchless gesture devices that provide assistive solutions for people who are hearing or physically impaired, there have been relatively few advances in gesture recognition technologies for people who are visually impaired. Research in this area has largely focused on improving the use of touchscreens for people with vision loss (Dim & Ren, 2014). One user study did identify an opportunity to solve a specific interaction challenge through GBI. Users in the study confirmed that a common activity for blind people is to listen to the TV, however, changing channels on the TV using a physical remote control is difficult when the buttons do not use braille (Dim & Ren, 2014). The research does not prototype or test a solution to this problem, but instead aimed to understand user preferences of creating their own or choosing from a list of predefined actions. Both research and industry developments for

gesture controlled devices that can address unique challenges for visually impaired people are limited at this time.

The examples of assistive technology in HCI research and commercially available products indicate how GBI can successfully address accessibility challenges. Although GBI is clearly addressing a need by creating more accessible ways to interact with technology, the use of these applications is limited to a narrow and specific population of people.

This review of gesture controlled devices in gaming, healthcare and assistive technology demonstrates that the reasons for adoption are diverse depending on the context in which they are required. GBI is solving problems where traditional tools have been less effective; however, the applications of this technology are limited to a handful of niche contexts. Microsoft's investment in creating the technology behind the Kinect suggests they had identified a unique opportunity to bring gesture-based interaction to the masses. Early reviews and the initial hype surrounding the Kinect hinted the technology was here to stay. The Kinect has proven to be popular—outside of home gaming—finding opportunities to solve specialized problems in healthcare and accessibility. What was originally intended to be used by a large audience has found its place in increasingly narrow applications. The specialized use cases seen throughout this review suggests GBI may not be appropriate for everyday use. Despite these learnings, research from the field of HCI offers a conflicting opinion. Research reviewed in the next chapter argues the true benefit of GBI is its natural and intuitive properties that make this technology easy to use and learn.

Is Theory at Odds with Practice?

Gesture-based interaction can offer a more ‘human’ and ‘natural’ way to interact with computers because it closely mirrors the way that people communicate with each other. This notion has greatly influenced the focus of research in the study and design of gesture technology. However, the body of research supporting this concept offers conflicting information on the definition and representation of ‘natural’ in technology.

Embodied Interaction

Researchers in the field of HCI have been discussing and applying the concept of *embodied interaction* for nearly two decades. Embodiment, as defined by Dourish (2001) is the “common way in which we encounter physical and social reality in the everyday world” (p. 100). Dourish’s *Where The Action Is* (2001) brings embodiment, as a philosophical concept rooted in phenomenology, into the context of HCI. Prior to Dourish, HCI emphasized cognitive processing to guide the research and design for useful systems, not surprising when you consider that computer scientists and psychologists worked together to build the foundations of HCI research (Preece et al., 2015). The focus of HCI on cognitive processing was criticized by others before Dourish (Winograd & Flores, 1986; Suchman, 1987; as cited in Preece et al., 2015); however, it was Dourish’s work that provided a new framework to help researchers consider how users create meaning and problem solve in real situations. The embodied interaction framework is particularly useful for gesture-based interactions as it highlights the importance of creating meaning through actions and engaging in skilled practice over disembodied or abstract interactions. Further, Dourish (2001) highlights the phenomenological emphasis on the natural in everyday activity. This emphasis on natural has significantly influenced HCI research, as can be seen in the following section.

Natural Interaction

Computers have commonly been considered both as an extension of self and as part of the physical world around us (Turkle, 1984). We personify computers by giving them human characteristics, viewing them symbolically as another person. We treat electronic devices as having functions necessary to survive, to the extent that they seem as if they are an appendage of our own body. We desire computers to fit seamlessly into our physical world, so much so that they seem to have been derived from nature itself. The use of the word ‘natural’ can encompass all the ways in which we desire computers to be represented in our lives. The use of the word ‘natural’ can also cause confusion for researchers and designers who strive to reflect this quality in human-computer interactions.

As Bestor (1993) points out, applying metaphors of nature to describe and market computers is nothing new. Descriptions of the early computer as an electronic ‘brain’ and describing networked systems as being susceptible to attacks from a ‘virus’ illustrates this point (Bestor, 1993). A review of HCI literature discussing natural user interfaces and gesture-based interaction suggests these methods can offer users a ‘natural’ or ‘intuitive’ computing experience. In fact, a review of decades of GBI literature point to ‘naturalness’ as a key benefit driving much of the research (Karam & Schraefel, 2005). However, Hansen and Dalsgaard (2015) criticize the overuse of the word ‘natural’ by marketers and researchers alike, stating the term is problematic not only because of its lack of clarity, but also because it neither represents or defines the underlying technology, nor does it exist in or from nature.

Naturalness, according to researchers, can be represented by both the interaction space and the interaction mode. Saffer (2008) and Preece et al. (2015) suggest the ‘naturalness’ of NUIs is indicative of the interaction mode. Interaction modes that more closely mirror human-to-

human communication, such as touch, gesture and speech, offer more natural experiences for humans as they communicate with computers. Saffer (2008) argues the artificiality of mechanical devices like a mouse or keyboard goes against human biology. His argument draws from a *Wired* interview with David Liddle, a professor of computer science at Stanford University and co-founder of Interval Research. Liddle theorizes that humans are born with physical skills that were originally designed for hunting and foraging, and so “why not try to make the tasks we do with our machines today look like the tasks the body was designed for?” (Bestor, 1993, para. 5). The small wrist motions required to navigate a mouse do not take advantage of the vast repertoire of human actions.

Preece et al. (2015) define a NUI as a system that “enables people to interact with a computer in the same way they interact with the physical world, through using their voice, hands and bodies” (p. 219). The type of input therefore predicates a more natural interface because it leverages everyday skills like touching an object, speaking and gesturing to others, rather than using a keyboard and mouse. It is not necessarily that learning to use a mouse or keyboard is unnatural and needlessly onerous, but leveraging existing talents can make systems easier to learn and reduce cognitive loads.

However, the work of Wigdor and Wixon (2011) further confuses our understanding of what makes systems more natural. They suggest that the word ‘natural’ is often understood by others as mimicking the real world (Wigdor & Wixon, 2011). An example of this can be seen in a common design practice called skeuomorphism that is used to familiarize new users with a GUI. Skeuomorphism in software references an object as it occurs in the real world (“Skeuomorphism is dead,” n.d.). A well-known example is the trash bin icon on your desktop used to discard unneeded files. By describing the function to designate files to be deleted as

trashing and designing an icon that looks like a trash bin, users can leverage their existing real world knowledge that unwanted items are placed in trash bins. Again, as in real life, users can ‘empty’ the computer’s trash bin (permanently deleting the files—like setting one’s trash on the curb for pick up). Wigdor and Wixon (2011) clarify that while mimicry is what is commonly (mis)understood as ‘natural,’ this concept is not what drives their design guidelines. They advise designers not to mimic the real world, instead, they suggest that one way to make users feel like a natural user is to “make an interface that mimics some other experience at which your user is already an expert” (Wigdor & Wixon, 2011, p. 13). They fail to clarify what this ‘other experience’ might be and how it differs from the user’s experiences in the real world. The author’s vague description of what is or is not mimicking the real world lacks examples and concrete instructions to help the reader understand their point of view.

While much of the HCI research often cites both ‘naturalness’ and ‘intuitiveness’ as clear benefits of NUIs and GBI, not only are these definitions vague and conflicting, the authors often fail to support these claims with any proven evidence. There is an apparent lack of information within HCI research on what it means to create ‘natural interactions’ and how to measure this quality in systems. The more generalized role of gesture to facilitate communication has been well researched in a number of interdisciplinary fields such as philosophy, psychology, anthropology, linguistics and cognitive science (Karam & Schraefel, 2005; van den Hoven & Mazalek, 2011). Yet researchers often fail to demonstrate exactly why and how gestures are useful in the context of HCI.

Grandhi, Joue and Mittelberg (2011) have also identified this gap in the literature, noting their work aims to take a small step towards identifying what makes gesture natural and intuitive in order to create guidelines for designers. Their work relies on findings from neuropsychology

to draw the following conclusions: dynamic gestures (movement through a range of motion) occurred 95% more often than static gestures (held with no movement); two-handed gestures occurred three times more often than one-handed gestures, and gestures that used a body part to hold an invisible tool occurred 75% more often than gestures that used a body part as the tool (Grandhi et al., 2011). For example, users pretended to hold a knife in one hand, an apple in the other hand and dynamically gestured as if they were cutting the apple. This type of research is significantly useful to a designer, however their research only addressed the type of gestures that feel more natural and intuitive and not why, overall, gestures create ‘natural’ experiences when communicating with computers.

Only a handful of authors have posed contrary opinions to the idea that GBI can inherently create ‘natural experiences’. In 2010, Don Norman authored the first notable journal article that took issue with research on the natural user interface. Norman (2010) proposed that NUIs were caught up in marketing rhetoric and had yet to be realized in well-designed products. Norman (2010) does advocate that NUIs and GBI present a positive future for computer interaction, stating:

gestural systems are indeed one of the important future paths for a more holistic, human interaction of people with technology. In many cases, they will enhance our feeling of control and empowerment, our convenience, and even our delight (pg. 10).

From this statement, it can be seen that Norman’s critique lies in existing NUI technology not fulfilling its goal of making NUIs natural and intuitive. Further, Norman argues that good design is independent of technology. Norman (2010) insists designers should return to the roots of interaction design in order to achieve truly intuitive systems, rather than attempting to merely recreate a GUI paradigm with new technology. O’hara, Harper, Mentis, Sellen and Taylor (2013)

assert that Norman's argument, while presenting an opposing opinion to NUIs, only superficially considers the execution of NUIs to be either natural or not natural, rather than taking issue with the vague notion of how to achieve 'naturalness' in designs.

In a similar line of argument to Norman's, Hansen and Dalsgaard (2015) are asking researchers and marketers to stop calling UIs natural. Their argument—which spends considerable time attempting to unpack the definition of natural—proposes the implications of this confusion threatens to undermine the integrity of HCI research. For marketers, applying the term 'natural' creates positive associations for purchasers. Really, who would want to buy a device that was advertised as 'unnatural' or 'artificial'? However, the term lacks the detail necessary to maintain the rigor of terminological precision in scholarly research (Hansen & Dalsgaard, 2015). Without precision, it makes ideas and arguments harder for future researchers to elaborate on or contest—an invaluable aspect of building a research community (Hansen & Dalsgaard, 2015). Neither Norman (2010), Hansen and Dalsgaard (2015), or O'hara et al. (2013) take issue with the value or impact of GBI and NUIs, but largely their criticism aims to force researchers to more clearly consider and define these fundamental features in an effort to better predict opportunities and challenges that may impact widescale adoption.

A critical review of the scholarly literature suggests the key benefits of implementing GBI is its ability to offer users a platform to perform simple, easy-to-learn, intuitive actions that can manipulate digital content or control devices. The use of meaningful, common gestures can create a more natural user experience and an interface so well-designed it can often feel invisible. There is a degree of expressiveness that is certainly not possible with a mouse. Similarly to Hansen and Dalsgaard (2015) and O'Hara et al. (2013), my intention is not to disprove the benefits of GBI and NUI, but to better understand how researchers are using the term to guide

the development of new systems. It is with this comprehension and clarity that we will be able to determine how best to apply the technology and ensure gesture controlled devices can move beyond the narrow current applications and more broadly support everyday computer tasks.

Theory Versus Practice

Despite the lack of clear definitions regarding the natural and intuitive qualities of GBI, we have seen how gesture technologies are solving problems for users where traditional tools (such as a mouse or touchscreen) cannot. The review of gesture controlled devices in gaming, healthcare and assistive technology demonstrated that the reasons behind adoption are diverse and dependent on the context in which they are seen as optimal or necessary in comparison with other forms of HCI. The adoption of GBI in healthcare environments is driven primarily by the importance of maintaining sterile environments through hands-free applications. In gaming, gesture-based applications encourage social connection through shared performative experience. Assistive technologies prioritize task efficiency over intuitive design. In all three fields, we see examples of how natural experiences and ease of use are not the driving factors behind the application and adoption of this technology, contrary to many of the claims made in the research in this field.

It is evident that the field of HCI lacks clarity on what is regarded as a natural experience in computer interaction; moreover, there is a disconnect between this popular theory and how the application of GBI is executed in practice. Again, my intention is not to question the value of natural experiences, but to draw attention to the lack of carryover into industry applications. If naturalness is not the primary reason for adoption in these applications, is it possible that there are other applications of GBI where natural experiences will matter and have the potential to increase adoption?

One possible application for gesture is within ubiquitous computing environments. Often referred to as “ubicomputing”, Weiser’s (1991) vision suggested a paradigm shift where computers would be embedded in everyday objects to create a network of computers that function without requiring user actions and as a result would often go unnoticed. Present day examples of ubicomputing systems are electronic toll systems, home automation and smart locks. The use of gesture in ubiquitous computing is not new, having been suggested by Weiser (1991), among others. Applications of GBI in ubiquitous computing environments could mean that gesture recognition technology will have the ability to sense implicit meaning from body language and facial expressions as compared to gestures that have explicit meaning and are performed in order to direct the computer to complete a specific task. This type of passive versus active interaction could present one avenue for gesture to become more mainstream; however, more powerful algorithms and machine learning would be required in combination with gesture recognition technology for this to become reality.

Chapters three and four have demonstrated that GBI can offer solutions and contributions to a number of interaction challenges. However, the introduction of this type of interaction method has also created several other challenges that have yet to be resolved. By analyzing the case studies and scholarly literature provided in this paper, it was evident that there are three areas that continue to present barriers to widespread adoption: technological, design and social.

A ‘NUI’ Set of Challenges

The introduction of a new technology often presents new challenges that may not have been considered or resolved at the time of release and gesture recognition technology is no exception. This chapter presents three areas that are presenting barriers to adoption. Each area consists of two main challenges that need to be addressed before GBI can take a more pronounced role in our everyday lives.

Technological Challenges

This paper discusses several gesture recognition products that are commercially available, such as the Kinect, LMC and Myo armband. The technological components that comprise these products function well, as technology development in this area has often focused on the ability of the system to recognize and translate gestures into commands. However, in keeping with the theme of this paper—to consider the implications of the technology from the perspective of the human, instead of the machine—I draw attention to two technological challenges that directly affect the person using the system.

One common challenge for the user is that the systems do not present ways to easily engage and disengage with the interface of these products. Without a “start and stop” feature it is difficult for the recognition system to understand when a gesture begins, ends, or what is not a meaningful action. Rosa and Elizondo (2014) describes this challenge as a ‘segmentation issue’ for the system to be able to identify when a gesture begins and ends. Consider if the system began triggering a task for the user when they were actually itching their ear or waving to try to get the attention of someone on the other side of the room. This difficulty can cause a cycle of unintended instructions to the system as a result of accidental gesturing and can lead to mounting frustrations for the user.

There are two companies that have attempted to address this issue. Thalmic Labs, the creator of the Myo, has included a specific gesture that unlocks the system for use. The two-fingered gesture requires users to ‘double tap’ their middle finger and thumb together. The set-up is well-intentioned, however not practical as it locks again within three seconds of recognizing the unlock gesture. The resulting limitation is that users can only complete one gesture at a time. The feedback from the system is designed to acknowledge that the system is unlocked by flashing a ‘ready’ icon on screen and sending a slight vibration to the armband. However, when the user waits until the system acknowledges the action, they have already missed the three second window to perform the next gesture and their intended task. Frustrations arise when the user needs to repetitively ‘unlock’ the system in order to accomplish any tasks.

Sesame, the creators of the smartphone that reads head gestures to assist people with physical disabilities, has combined speech with gesture to address this problem. The software, which uses a Google Android smartphone, requires users to say “open sesame” before each gesture-based interaction (“Sesame,” n.d.). This type of activation is common in speech-based interfaces such as Amazon Alexa and Google Home. The integration of speech activation with gestured actions presents an interesting combination. One that could solve this problem for home or work environments, but may not work well in loud or busy public spaces.

In addition to the challenges of activating and deactivating the recognition system, technologies are limited by interaction zones. In order to interact with camera-based gesture systems, users must place themselves in specific areas, known as interaction zones—where the system can recognize and process their gestures. The need to be in a specific physical space to interact with the system is a limitation in and of itself. For example, the Kinect interaction zone is approximately 0.4 meters in ‘near mode’ and 3-4 meters in ‘normal mode’ (“KinectInteraction

Concepts,” n.d.). This area is the space where the Kinect camera projects 3D mapping and can recognize movements within the designated space. The expansive size of the interaction zone for the Kinect is useful for gaming and interactive installations intended to be used by multiple users, as demonstrated by Design I/O’s Connected Worlds exhibit. Interaction zones must be tailored to the intended use of the technology. For instance, an expansive interaction zone would be less useful for doctors in the operating room because they would need to walk in and out of the zone to conduct their gestures, essentially disrupting their workflow. The LMC offers a small interaction zone of approximately 0.23 m³ making it easier for users to remove their hands from the zone without having to move their whole body (Rosa & Elizondo, 2014). Different sized interaction zones can be beneficial for different applications, but this restriction limits the mobility of the user and further restricts how, where and when the system can be used effectively. These technological issues present limitations for integrating GBI into home and work tasks.

Design Challenges

Many companies have created gesture recognition systems with broad applications in mind. For example, both the Kinect and LMC are commercial products designed for game and home use (Mewes et al., 2017). To find new applications/uses of their technology, these companies have turned to software development kits (SDKs) as a way to empower others to explore and build new applications for their existing hardware. As mentioned earlier, this was an important element of the success of the Kinect, whose rise to popularity relied mostly on the creative applications of the technology outside of the gaming world. The LMC and Myo have also created an SDK and the resulting applications created by developers can be added to and

explored through their respective marketplaces (“Leap Motion,” n.d., “Myo Gesture Control Armband,” n.d.). Further, Google’s Project Soli—a gesture recognition chip so small it can be embedded into wearable devices—has announced they will be releasing an SDK with a call to developers to build a community to test and evolve their project through the creation of new applications (“Project Soli,” n.d.). In many cases, the companies behind gesture recognition tools rely on SDKs and the communities they build to find new and creative ways to use their technology.

While there is no shortage of technological innovations in the field of gesture recognition, the problem with their design is that they are not conceived to solve any one particular problem for the user. Mewes et al. (2017) argue that custom-built, task-specific products can improve the robustness and usability of gesture-based devices. The reliance on developers via SDKs to create new uses of this technology represents a tremendous step forward in the proliferation of gesture recognition systems. But the commonality between the new applications created by SDKs is that they often seek to simply replace standard mouse and keyboard activities with gesture.

Designers need to think beyond improving common keyboard/mouse interactions and more broadly consider the tasks that people need to accomplish. Commercially available speech-based devices have done this well. Amazon’s Alexa and the Google Home speech interfaces integrate with desktop and smartphone applications when needed, but they have been designed to assist users with tasks that eliminate the need for screen-based devices. For example, Google Home can help you search for a recipe and read the recipe aloud step-by-step, repeating instructions as needed without relying on screens. A demo video from Thalmic Labs shows a person in the kitchen wearing a Myo armband and using the tool to follow a recipe without needing to touch their tablet with their dirty hands. Indeed, the latter example is helpful but the

solution simply replaces one interaction mode with another. Designers of gesture-based systems need to capture the task-specific focus that designers of speech-based interfaces have adopted. Instead of trying to translate existing interactions into gesture-based interactions, designers should consider what the ultimate goal of the user is and then design specific products to address that need.

There is also a lack of standardized gesture vocabularies among creators of gesture recognition devices. This lack of a universal ‘language’ can be confusing for users who develop the necessary skills to use one system and then have to learn a whole new system in another context/environment. Continuity between applications in a system is important, but so is continuity across multiple systems. There is no expectation for all systems to be identical, as this would defeat the purpose of having multiple products on the market, however there should be commonalities to ensure users can easily work within multiple systems. For example, computer actions across all desktop computing systems are unique, but still present similarities that make transitioning from one platform to the other much easier. The ‘cut and paste’ keyboard shortcut is a great example of this in desktop computing as it varies only slightly between Microsoft and Apple. Touch gestures for iPhone and Android smartphones and tablets, such as the two-finger ‘pinch’ to zoom in and out, have also been created consistently across applications.

The pinch to zoom gesture was popularized on Apple’s iPhone and later adopted by competitors like Samsung and Google. Apple provides gesture guidelines for developers who are creating apps to be used on Apple devices. Their guidelines note the importance of having standards for users who expect actions to mean the same thing across all applications on the Apple system (“iOS Human Interface Guidelines,” n.d.). Yet, their history of patent claims and courtroom battles (Patel, 2012) suggest Apple feels strongly that competitors should not be able

to leverage proprietary interaction languages. Apple's legal stance on the issue aside, the proliferation of similar gestures across other smartphones shows that organizations are working towards creating vocabularies that share more similarities than differences.

There is a wide variety of interaction languages among existing touchless gesture systems. This could be explained by the fact that designers may be purposely ignoring convention in order to be innovative (Norman & Nielsen, 2010). The apparent lack of continuity between these systems could be a result of the limitations of each particular device (Grandhi et al., 2011). Alternatively, it could be a reflection of the infancy of gesture-controlled devices. Regardless of why so many interaction languages have been created, the current offerings of touchless gesture devices can be frustrating for the user. Currently, when a user interacts with new systems, they are required to learn new sets of gestures which may conflict with their existing learned 'vocabulary'. The conflict between languages essentially eliminates a user's intuition on how to best interact with new systems (Grandhi et al., 2011). Rather than create new, more complicated gestures, designers should focus on the creation of a common, simplified language across gesture-based systems to improve the user experience. It would benefit users if gesture languages were similar, as we have seen with desktop and smartphones systems. Resolving these design challenges can help GBI find additional use cases and allow users to learn new systems more easily.

Social Challenges

The application and adoption of GBI have largely occupied private settings such as the home or operating theatre. In a culture where mobile computing has become ubiquitous due to the widespread adoption of smartphones and tablets, designers and researchers need to consider

how GBI might change the way people use technology in a variety of settings. Rico and Brewster (2010) assert that GBI faces adoption issues because it forces individuals to confront a whole new standard of social acceptability. The use of technologies in the public sphere can place a barrier between two groups of people—those using the technology (users) and those who are not (spectators). An example of this can be seen with the introduction of Google Glass in 2013 (Gross, 2014). Glass is one of many examples of how new, unfamiliar technology can create uncomfortable social interactions between technology users and spectators.

GBI requires users to adopt behaviours that can be considered embarrassing or disruptive when observed out of context. Consider, for example, waving your hand at a virtual screen to activate it or mimicking the gesture of turning a key in a lock to deactivate it. As a user, you might be self-conscious of the action you are performing, wondering if others are passing judgement on you. An ‘in the wild’ study evaluating the social acceptability of gestures in a variety of settings confirmed that some users felt self-conscious or simply refused to perform the gesture in public settings (Williamson, Brewster, & Vennelakanti, 2013). Likewise, as a spectator, seeing someone else use this gesture, you might question what they are doing and why. To illustrate this point, consider when Bluetooth technology was first introduced. Early adopters purchased Bluetooth headsets to wirelessly receive phone calls right into a small earpiece. Anyone standing within a few meters of the user receiving a call often felt bewildered wondering who, if anyone, the user was talking to. Continued exposure for spectators helped Bluetooth headsets move past this perplexing stage and become widely accepted as a hands-free tool for talking on the phone.

The initial confusion caused by the misinterpretation of actions performed by someone in your physical space can be unnerving for spectators. This notion is further supported by the work

of Montero, Marshall and Subramanian (2010) which separates social acceptability into two elements: the user's social acceptance and the spectator's social acceptance. Social acceptance for the user is reflected by the overall experience of using the technology to complete a task. For example, was the user comfortable performing the gesture? Did they feel awkward? Were they embarrassed? Users must weigh the risks of looking strange with the perceived benefits of the system to determine if they are willing to use it.

Conversely, the spectator's social acceptance relates directly to the positive or negative impression of the user's actions (Montero et al., 2010). To further understand these impressions, consider if the spectator(s) perceived the user's actions to be strange. Or, if the spectator understood what the user was doing. For the spectator, both the user's actions and the resulting effect of the action contributes to their understanding of the gesture, and therefore the social acceptance of the gesture (Montero et al., 2016). The results of a user study conducted by Montero et al. (2016) showed that gestures where the action is revealed but the effect is concealed are unacceptable to spectators. An example of this 'suspenseful' classification of gestures is if the user wrote a large 'X' in the air to set a mobile phone to silent mode (Montero et al., 2016). Spectators are able to see the user drawing in the air, but they cannot tell what effect the action had on the device. Other gestures where the action is concealed but the effect is revealed or where both the action and effect is either hidden or revealed were more socially acceptable. The authors draw on the theory of embodied interaction (Dourish, 2001) to suggest that actions embodying meaning are important for both the user and the spectator. If the spectator cannot see or understand the results, there is a disconnect while trying to comprehend the true meaning of the gesture. The implications of this study do not suggest that spectators need to know every detail of every gesture, but rather a general notion in order to connect action and

effect. This presents a unique challenge for designers who will need to create gestures that meet the needs of both users and spectators.

The public relationship between user and spectator further becomes an issue when the user's privacy is threatened. As a result of their 'out in the open' style of interaction, gestures are susceptible to invasions of privacy. Privacy and security concerns are not unique to GBI, rather most new technologies face similar adoption challenges. A common privacy concern for users is the collection and misuse of personal data by organizations. As a result, it is up to the organization collecting the information to be transparent on how personal data will be stored and used. The creation of clear and easy to read privacy policies or terms of use can quell fears from users who are hesitant to pass over their personal information in exchange for access to the technology. In addition to personal data concerns, creators of gesture-controlled devices need to be cognizant of possible privacy invasions from spectators. Designing smaller interaction zones for gesture recognition and small, subtle gestures may reduce initial concerns. If product designers can work towards addressing these challenges that become apparent when GBI is used in public settings, gesture has a better chance of being socially accepted and more widely adopted.

Conclusion

The nature and extent of the adoption of gesture-based interfaces in our everyday life are largely dependent upon the ability of product designers to address these identified technological, design and social barriers. Researchers in HCI studies are advocating that the key benefit of implementing GBI is its ability to offer users a platform to perform simple, easy-to-learn, intuitive actions to control devices. The research also points to the potential of meaningful gestures, ones that mirror human-to-human communication. In theory, meaningful gestures can result in experiences that are deemed more natural by users than interacting with a computer through a mouse or keyboard.

The reality is that gesture-based interactions, as they currently exist, only solve specific problems in specific contexts. The technology can increase shared connections in gaming, maintain sterile environments in the operating room and address accessibility issues for people with auditory or physical impairments. Despite these incredible applications of this technology, the widespread adoption of gesture-based interaction remains limited at this time.

In order to realize a future where gesture is as common as touchscreens are today, theorists and researchers need to establish an agreed upon definition of 'natural' when used in the context of gesture-based interaction. Without a mutual understanding of this term, it is hard for designers to create systems that achieve this goal, let alone measure their effectiveness. Further to this, product creators need to work towards resolving the technological, design and social challenges that are presently barring the widespread adoption of this technology. While addressing any one of these issues is no small feat, it is my hope that researchers and designers in the not-so-distant future will be able to work together to overcome these challenges and contribute to the widespread adoption and integration of gesture in our everyday lives. Gesture

can empower us to learn in new, creative ways and help us work more efficiently. Above all of these possibilities, and perhaps most importantly, this technology can protect my smartphone from being ruined by my messy hands while I am cooking. What a wonderful world that would be.

References

- Bansal, B. (2016). Gesture recognition: A survey. *Gesture*, 139(2).
- Bestor, T. (1993, May 1). Dogs don't do math. *WIRED*. Retrieved from <https://www.wired.com/1993/05/dogs/>
- Carroll, J. M. (n.d.). The Encyclopedia of Human-Computer Interaction, 2nd Ed. Retrieved July 13, 2017, from <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro>
- Connected worlds. (n.d.). Retrieved July 15, 2017, from <http://design-io.com/projects/ConnectedWorlds/>
- Dim, N. K., & Ren, X. (2014). Designing motion gesture interfaces in mobile phones for blind people. *Journal of Computer Science and Technology*, 29(5), 812–824. <https://doi.org/10.1007/s11390-014-1470-5>
- Dourish, P. (2001). *Where the action is: The foundations of embodied interaction*. The MIT Press.
- GestSure. (n.d.). Retrieved August 2, 2017, from <https://www.gestsure.com/>
- Giles, J. (2010). Inside the race to hack the Kinect. *New Scientist*, 208(2789), 22–23. [https://doi.org/10.1016/S0262-4079\(10\)62989-2](https://doi.org/10.1016/S0262-4079(10)62989-2)
- Grandhi, S. A., Joue, G., & Mittelberg, I. (2011). Understanding naturalness and intuitiveness in gesture production: Insights for touchless gestural interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 821–824). ACM. <https://doi.org/10.1145/1978942.1979061>
- Gross, A. (2014, March 4). What's the Problem with Google Glass? *The New Yorker*. Retrieved from <http://www.newyorker.com/business/currency/whats-the-problem-with-google-glass>

- Hansen, L. K., & Dalsgaard, P. (2015). Note to self: Stop calling interfaces natural. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives* (pp. 65–68). Aarhus University Press. <https://doi.org/10.1017/CBO9781107415324.004>
- Hettig, J., Saalfeld, P., Luz, M., Becker, M., Skalej, M., & Hansen, C. (2017). Comparison of gesture and conventional interaction techniques for interventional neuroradiology. *International Journal of Computer Assisted Radiology and Surgery*, 1–11. <https://doi.org/10.1007/s11548-017-1523-7>
- Hruska, J. (2014, June 9). Microsoft begs devs to pay attention to Kinect 2 for Windows, even after Xbox One de-bundling. *Extreme Tech*. Retrieved from <https://www.extremetech.com/gaming/183913-microsoft-begs-devs-to-pay-attention-to-kinect-2-for-windows-even-after-xbox-one-de-bundling>
- iOS Human Interface Guidelines. (n.d.). Retrieved October 23, 2016, from <https://developer.apple.com/ios/human-interface-guidelines/visual-design/animation/>
- Karam, M., & Schraefel, M. (2005). A taxonomy of gestures in human computer interactions. *Computing Service*.
- Kessler, S. (2011, March 9). Microsoft Kinect sales top 10 million, set new Guinness World Record. *Mashable*. Retrieved from <http://mashable.com/2011/03/09/kinect-10-million/#E72sEDYAf5qq>
- KinectInteraction Concepts. (n.d.). Retrieved August 21, 2017, from <https://msdn.microsoft.com/en-us/library/dn188673.aspx>
- Kondori, F. A., Yousefi, S., Liu, L., & Li, H. (2014). Head operated electric wheelchair. In *2014 Southwest Symposium on Image Analysis and Interpretation* (pp. 53–56). <https://doi.org/10.1109/SSIAI.2014.6806027>

- Leap Motion. (n.d.). Retrieved July 26, 2017, from <https://www.leapmotion.com/>
- Mapari, R. B., & Kharat, G. (2016). American static signs recognition using leap motion sensor. In *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies* (p. 67:1-67:5). New York, NY, USA: ACM.
<https://doi.org/10.1145/2905055.2905125>
- Mewes, A., Hensen, B., Wacker, F., & Hansen, C. (2017). Touchless interaction with software in interventional radiology and surgery: A systematic literature review. *International Journal of Computer Assisted Radiology and Surgery*, 12(2), 291–305.
<https://doi.org/10.1007/s11548-016-1480-6>
- Montero, C. S., Marshall, M. T., & Subramanian, S. (2010). Would you do that?: Understanding social acceptance of gestural interfaces. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 275–278). ACM.
- MotionSavvy. (n.d.). Retrieved July 26, 2017, from <http://www.motionsavvy.com/index.html>
- Myers, B. A. (1998). A brief history of human-computer interaction. *Interactions*, 5(2), 44–54.
<https://doi.org/10.1145/274430.274436>
- Myo Gesture Control Armband. (n.d.). Retrieved August 5, 2017, from <https://www.myo.com/>
- Norman, D. A. (2010). Natural user interfaces are not natural. *Interactions*, 17(3), 6–10.
<https://doi.org/10.1145/1744161.1744163>
- Norman, D. A., & Nielsen, J. (2010). Gestural interfaces: A step backward In usability. *Interactions*, 17(5), 46–49. <https://doi.org/10.1145/1836216.1836228>
- nz technologies inc. (n.d.). Retrieved August 2, 2017, from <http://nztech.ca/>
- O'hara, K., Harper, R., Mentis, H., Sellen, A., & Taylor, A. (2013). On the naturalness of touchless: Putting the “interaction” back into NUI. *ACM Transactions on Computer-Human*

- Interaction*, 20(1), 1–25. <https://doi.org/10.1145/2442106.2442111>
- Patel, N. (2012, August 30). The myth of pinch-to-zoom: how a confused media gave Apple something it doesn't own. *The Verge*. Retrieved from <https://www.theverge.com/2012/8/30/3279628/apple-pinch-to-zoom-patent-myth>
- Peckham, M. (2011, January 21). Top 15 Kinect hacks (so far). *PCWorld*. Retrieved from http://www.pcworld.com/article/217283/15_radical_kinect_hacks.html
- Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design: Beyond human-computer interaction* (Fourth ed.). John Wiley & Sons Ltd.
- Project Soli. (n.d.). Retrieved August 5, 2017, from <https://atap.google.com/soli/>
- Rico, J., & Brewster, S. (2010). Usable gestures for mobile interfaces: Evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 887–896). ACM. <https://doi.org/10.1145/1753326.1753458>
- Rosa, G. M., & Elizondo, M. L. (2014). Use of a gesture user interface as a touchless image navigation system in dental surgery: Case series report. *Imaging Science in Dentistry*, 44(2), 155–160. <https://doi.org/10.5624/isd.2014.44.2.155>
- Saffer, D. (2008). *Designing gestural interfaces (First Ed.)*. O'Reilly Media.
- Schwarz, L. A., Bigdelou, A., & Navab, N. (2011). Learning gestures for customizable human-computer interaction in the operating room. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 129–136).
- Scopis. (n.d.). Retrieved August 2, 2017, from <https://navigation.scopis.com/>
- Sesame. (n.d.). Retrieved August 1, 2017, from <http://www.sesame-enable.com/>
- Skeuomorphism is dead, long live skeuomorphism [Blog post]. (n.d.). 2017. Retrieved from <https://www.interaction-design.org/literature/article/skeuomorphism-is-dead-long-live->

skeuomorphism

- Soutschek, S., Penne, J., Hornegger, J., & Kornhuber, J. (2008). 3-D gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1–6). IEEE. <https://doi.org/10.1109/CVPRW.2008.4563162>
- Spielberg, S. (2002). *Minority Report [Motion picture]*. Twentieth Century Fox Film Corporation.
- TedCas. (n.d.). Retrieved August 2, 2017, from <http://www.tedcas.com/>
- The open Kinect project – THE OK PRIZE – get \$3,000 bounty for Kinect for Xbox 360 open source drivers [Blog post]. (2010, November 4). Retrieved from <https://blog.adafruit.com/2010/11/04/the-open-kinect-project-the-ok-prize-get-1000-bounty-for-kinect-for-xbox-360-open-source-drivers/>
- Therapixel. (n.d.). Retrieved August 2, 2017, from <http://www.therapixel.com/>
- Turkle, S. (1984). *The second self: Computers and the human spirit*. Simon and Schuster.
- van den Hoven, E., & Mazalek, A. (2011). Grasping gestures: Gesturing with physical artifacts. *AI EDAM*, 25(3), 255–271. <https://doi.org/10.1017/S0890060411000072>
- Wachs, J. P., Stern, H. I., Edan, Y., Gillam, M., Handler, J., Feied, C., & Smith, M. (2008). A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association : JAMIA*. United States. <https://doi.org/10.1197/jamia.M241>
- Weinberger, M. (2015, September 8). The downfall of Kinect: Why Microsoft gave up on its most promising product. *Business Insider*. Retrieved from <http://www.businessinsider.com/why-microsoft-xbox-kinect-didnt-take-off-2015-9>

- Weiser, M. (1991, September). The computer for the 21st century. *Scientific American*, 94–104.
- Wigdor, D., & Wixon, D. (2011). *Brave NUI world: Designing natural user interfaces for touch and gesture*. Morgan Kaufmann.
- Williamson, J. R., Brewster, S., & Vennelakanti, R. (2013). Mo!Games : Evaluating mobile gestures in the wild. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (pp. 173–180). ACM. <https://doi.org/10.1145/2522848.2522874>
- Wipfli, R., Dubois-Ferrière, V., Budry, S., Hoffmeyer, P., & Lovis, C. (2016). Gesture-controlled image management for operating room: A randomized crossover study to compare interaction using gestures, mouse, and third person relaying. *PloS One*, 11(4), e0153596. <https://doi.org/10.1371/journal.pone.0153596>