IMAGINARY INTERFACES

SEAN GUSTAFSON

Dissertation submitted for the degree of

Dr. rer. nat.

Supervised by
Prof. Dr. Patrick Baudisch

Hasso Plattner Institute
Potsdam, Germany
June 2013

ABSTRACT

The size of a mobile device is primarily determined by the size of the touchscreen. As such, researchers have found that the way to achieve ultimate mobility is to abandon the screen altogether. These wearable devices are operated using hand gestures, voice commands or a small number of physical buttons. By abandoning the screen these devices also abandon the currently dominant spatial interaction style (such as tapping on buttons), because, seemingly, there is nothing to tap on. Unfortunately this design prevents users from transferring their learned interaction knowledge gained from traditional touchscreen-based devices.

In this dissertation, I present *Imaginary Interfaces*, which return spatial interaction to screenless mobile devices. With these interfaces, users point and draw in the empty space in front of them or on the palm of their hands. While they cannot see the results of their interaction, they obtain some visual and tactile feedback by watching and feeling their hands interact. After introducing the concept of Imaginary Interfaces, I present two hardware prototypes that showcase two different forms of interaction with an imaginary interface, each with its own advantages: mid-air imaginary interfaces can be large and expressive, while palm-based imaginary interfaces offer an abundance of tactile features that encourage learning.

Given that imaginary interfaces offer no visual output, one of the key challenges is to enable users to discover the interface's layout. This dissertation offers three main solutions: offline learning with coordinates, browsing with audio feedback and learning by transfer. The latter I demonstrate with the *Imaginary Phone*, a palm-based imaginary interface that mimics the layout of a physical mobile phone that users are already familiar with.

Although these designs enable interaction with Imaginary Interfaces, they tell us little about why this interaction is possible. In the final part of this dissertation, I present an exploration into which human perceptual abilities are used when interacting with a palm-based imaginary interface and how much each accounts for performance with the interface. These findings deepen our understanding of Imaginary Interfaces and suggest that palm-based imaginary interfaces can enable stand-alone eyes-free use for many applications, including interfaces for visually impaired users.

ZUSAMMENFASSUNG

Die Größe mobiler Geräte ist vornehmlich bestimmt durch die Größe des Berührungsbildschirms. Forscher haben daher erkannt, dass der Weg zur äußersten Mobilität in der kompletten Aufgabe des Bildschirms liegt. Solche tragbaren Geräte werden durch Handgesten, Sprachbefehle oder eine kleine Anzahl physikalischer Tasten gesteuert. Mit der Aufgabe des Bildschirms geben diese Geräte allerdings auch den momentan weitverbreiteten Stil räumlicher Interaktion auf (zum Beispiel das Betätigen von Tasten), da scheinbar nichts existiert, das man betätigen kann. Leider verhindert diese Entwicklung, dass Benutzer Interaktionswissen, welches sie sich auf herkömmlichen berührungsempflindlichen Geräten angeeignet haben, anwenden können.

In dieser Doktorarbeit stelle ich *Imaginary Interfaces* vor, imaginäre Benutzerschnittstellen, die räumliche Interaktionen auf bildschirmlosen mobilen Geräten ermöglichen. Diese Schnittstellen erlauben Benutzern, im leeren Raum vor ihnen oder auf ihren Handfläche zu zeigen und zu zeichnen. Zwar können Benutzer die Ergebnisse ihrer Interaktion nicht sehen, sie erhalten jedoch visuelle und taktile Rückmeldung dadurch, dass sie ihre Hände während der Interaktion beobachten und fühlen. Nach der Einführung des Imaginary Interfaces Konzepts stelle ich zwei Hardware-Prototypen vor, die zwei verschiedene Arten von Interaktionen mit Imaginary Interfaces demonstrieren, jeweils mit ihren eigenen Vorteilen: Imaginary Interfaces in der Luft können groß und ausdrucksstark sein, während Imaginary Interfaces basierend auf Handflächen eine Fülle von taktilen Merkmalen aufweisen, die das Erlernen unterstützen.

Die fehlende visuelle Ausgabe führt zu einer der Hauptherausforderungen von Imaginary Interfaces, nämlich Benutzern zu ermöglichen, die Anordnung der Benutzerschnittstellen herauszufinden. Diese Doktorarbeit stellt drei Lösungen vor: vorheriges Lernen mit Koordinaten, Durchsuchen mit Tonrückmeldung und Lernen durch Transfer. Letztere demonstriere ich mit *Imaginary Phone*, einem Imaginary Interface basierend auf Handflächen, das die den Benutzern schon vertraute Anordnung eines physikalischen Mobiltelefons imitiert.

Obwohl diese Lösungen die Interaktion mit Imaginary Interfaces ermöglichen, können sie keine Aussage darüber treffen, warum eine solche Interaktion möglich ist. Im letzten Teil dieser Doktorarbeit untersuche ich, welche menschlichen Wahrnehmungsfähigkeiten während der Interaktion mit Imaginary Interface basierend auf

Handflächen genutzt werden und zu welchem Ausmaß jede dieser Wahrnehmungsfähigkeiten zur Effizienz bei der Benutzung beiträgt. Diese Ergebnisse vertiefen unser Verständnis von Imaginary Interfaces und legen nahe, dass Imaginary Interfaces basierend auf Handflächen die eigenständige und blickfreie Benutzung von vielen Anwendungen ermöglichen können, eingeschlossen Benutzerschnittstellen für sehbehinderte Benutzer.

PUBLICATIONS

Much of the research in this dissertation was previously published in the following articles. Specific chapters and sections that directly derived from each publication are listed:

Gustafson, S., Bierwirth, D. and Baudisch, P. (2010). Imaginary Interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of UIST*, pages 3–12.

- Chapter 3
- Section 4.1

Gustafson, S., Holz, C. and Baudisch, P. (2011). Imaginary Phone: learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of UIST*, pages 283–292.

- Chapter 5. Except Section 5.4, which contains a previously unpublished study.
- Section 4.2 and Section 4.3

Gustafson, S. (2012). Imaginary Interfaces: touchscreen-like interaction without the screen. In *Extended Abstracts of CHI (Doctoral Consortium)*, pages 927–930.

Gustafson, S., Rabe, B. and Baudisch, P. (2013). Understanding palm-based imaginary interfaces: the role of visual and tactile cues when browsing. In *Proceedings of CHI*, pages 889–898.

- Chapter 6

CONVENTIONS

USE OF 'I' VS. 'WE'

In the introductory and concluding matter of this dissertation I adopt the convention of referring to Imaginary Interfaces as *my* work. I do this to maintain consistency despite the fact that the concept of Imaginary Interfaces and the ideas contained in this dissertation were developed together with my doctoral advisor, Prof. Dr. Patrick Baudisch, without whom none of the ideas in this dissertation would have been realized.

Furthermore, various sections of this dissertation originated from work produced in concert with my collaborators. In the sections derived from these publications I adopt the convention of using the first person plural, *we*, when referring to those who conducted the research (for example, "We investigated…") but retain the first person singular, *I*, when referring to the narrative of the dissertation itself (for example "As I will show in Chapter 5…").

CAPITALIZATION OF IMAGINARY INTERFACES

Throughout this dissertation I adopt the convention of capitalizing Imaginary Interfaces when referring to the *concept* of Imaginary Interfaces or to the title of the research project (for example, "I begin the exploration of Imaginary Interfaces by…"). I use lower case to refer to specific instantiations of imaginary interfaces (for example, "Users interacted with an imaginary interface…").

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my tireless advisor Prof. Dr. Patrick Baudisch for the huge amount of feedback, generous financial support and for giving me this opportunity. I could never account for all that you taught me. Thank you sincerely.

Thanks to my internship advisor Dr. Ruth Rosenholtz and my other collaborators: Daniel Bierwirth, Christian Holz, Bernhard Rabe, Christian Loclair, Christian Steins, Prof. Dr. Sebastian Boring, Dr. Dominikus Baur, Prof. Dr. Andreas Butz, Prof. Dr. Stéphane Huot, Prof. Dr. Wendy Mackay, Dr. Julie Wagner, Dr. Mathieu Nancel. I learned so much working with you.

Thanks to all my labmates: Christian Holz, Henning Pohl, Dr. Anne Roudaut, Pedro Lopes, Stefanie Müller, Dr. Liwei Chan, Lung-Pan Cheng and Dr. Dominik Schmidt. Thanks for the last years of brainstorming, critique, proofreading and encouragement. Also thank you to the many bachelor students, masters students and interns who helped by participating in my studies, engaging in brainstorming sessions, modeling for photos and much more. And thanks also to the many people from the community for participating in my experiments. Thanks for taking the time to come down and for enduring the occasionally monotonous study tasks.

Very special thanks to Prof. Dr. Hasso Plattner for funding this wonderful institute. You have created the perfect environment for scholarship and research that allowed me to wholly concentrate on my work. I cannot thank you enough.

Thank you to Dr. Matthias Ringwald for providing his lib-hidsupport software that enabled touch event injection on the iPhone and to Jochen Penne and Prof. Dr. Thorsten Ringbeck from PMDTechnologies GmbH for advice and access to depth sensing camera equipment.

Thanks to Dr. David Dearman and Dr. Esben Pederson for reviewing early versions of my CHI 2013 paper and providing excellent feedback. Also, I would like the thank the anonymous paper reviewers of my submitted papers and to Prof. Dr. Per Ola Kristensson and Prof. Dr. Stephen Brewster, my doctoral consortium mentors. Your extensive feedback was essential for me to grow as a researcher. Thanks to my external dissertation reviewers Prof. Dr. Michael Rohs and Prof. Dr. Kasper Hornbæk, whose feedback was also instrumental in wrapping up four years of work.

Thank you again to Christian Holz for proofreading this dissertation and for guiding me through the administrative process and to Dr. Dominik Schmidt for translating the abstract into German.

Thank you to my mentors Prof. Dr. Pourang Irani and Rick Duff, whom I could always turn to for advice or encouragement. Thank you.

Special thanks to my old friends back in Canada and to my new friends scattered around the world. You were always available for support and for distraction. Thanks for being there.

Huge thanks to my lovely girlfriend Hélène, who put up with so much over these last few years and never stopped encouraging and supporting me, especially when I needed it the most.

Most importantly I would like to thank my family, especially my mother. I am in debt to you all for providing a loving refuge on the other side of the planet, for instilling in me the values of respect, patience and hard work, and for putting up with me when I was too busy to call.

Finally, a special thanks to my late father, who passed away while I was pursuing this degree. I wish I could have shared this accomplishment with you but I will always cherish the time we had together. Thank you for the opportunities you gave me and for never doubting I could accomplish anything I wished. I love you Dad.

CONTENTS

ABBREVIATIONS

ANOVA	analysis of variance – statistical hypothesis test
AR	augmented reality
DSLR	digital single lens reflex – style of digital camera
GUI	graphical user interface
IR	infrared
LED	light emitting diode
M	mean
NUI	natural user interface
SD	standard deviation
TUIO	tangible user interface objects – protocol for sending touch positions between applications
UI	user interface
VR	virtual reality
WLAN	wireless local area network – also known as WiFi

1 INTRODUCTION

The predominant interaction style for modern computer systems is based on a direct and spatial metaphor: user interface (UI) elements are located at specific locations and in order to interact with an element (e.g., to select a button) users point at its location and perform an action to complete the selection (e.g., click the mouse or tap the screen). This is true of most modern computer interfaces, such as desktops, laptops, large interactive displays, tablets and mobile phones.

For mobile devices, the size of the screen defines the size of the device. Therefore, for some applications where mobility is important, researchers realized the only way to achieve ultimate mobility is to abandon the screen altogether (Ni and Baudisch 2009). These wearable devices must resort to non-spatial interaction styles, such as voice commands or hand gestures, because there is simply no screen to tap on. For instance, Gesture Pendant (Starner et al. 2000) allows users to perform a series of commands out of a finite gesture vocabulary, such as “open door” or “lower blinds”. Interfaces like this do not provide spatial interaction but instead employ *categorical* gestures, where each gesture invokes a single operation. This unfortunately requires the user to learn a large set of gestures, which makes these interfaces unwieldy for general purpose interaction. Consequently, for all their gains in mobility, they severely limited the learnability of the device by discarding the spatial interaction style that users are already familiar with.

In this dissertation, I address this problem by introducing *Imaginary Interfaces*, spatial non-visual interfaces for mobile devices that users interact with by pointing using their dominant hand either in the empty space in front of them or on their non-dominant hand. Imaginary Interfaces achieve *ultra-mobility* by eschewing all visual feedback from the device and moving the interaction away from the device itself into the user’s environment. This allows the size of the device to be limited by the sensing element (in this case, a chest-mounted camera) and not by the size of the interaction surface, while retaining the spatial interaction style of other modern interfaces.

To use an imaginary interface, the user must *imagine* the interface. They build up an understanding of the interface as they draw on an imaginary interaction plane or transfer knowledge of the interface from their experience with a corresponding physical device. An example of these two interactions are shown Figure 1.1. The first, in Figure 1.1a, is the simplest version of Imaginary Interfaces and the first I present in this

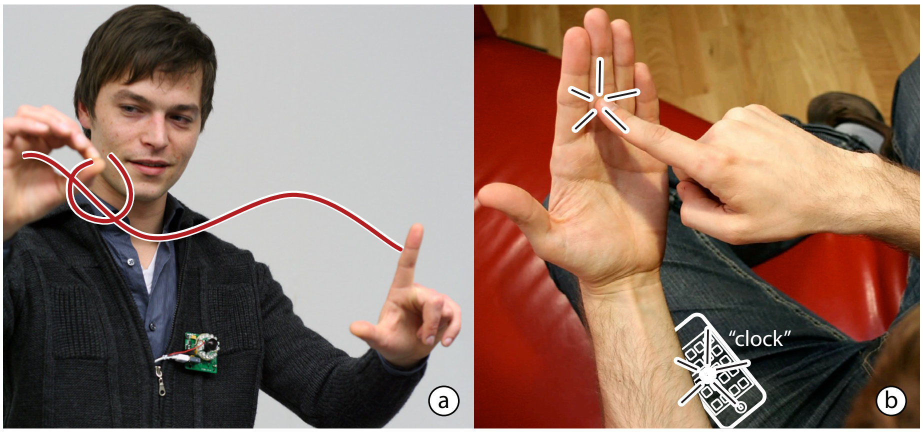


Figure 1.1: With Imaginary Interfaces, user's interact with an interface that exists in their *imagination*, either (a) in mid-air or (b) on the palm of their hand.

dissertation. In this example, the user draws a stock curve in mid-air using his imaginary interface while communicating with his stockbroker. The second example shown in Figure 1.1b is the *Imaginary Phone*, an imaginary interface that can be used as a shortcut for a mobile phone by allowing users to *transfer* the spatial knowledge they have from an existing physical device to an imaginary interface.

Although the lack of visual feedback limits what is possible with Imaginary Interfaces (e.g., the user cannot watch a movie), as I will argue in this dissertation, there are many situations where visual feedback is not strictly required and where Imaginary Interfaces could prove useful.

The thesis statement of this dissertation is therefore as follows:

> *By exploiting users' visuospatial memory, by using their palm as an interaction surface, and by supplying audio feedback, it is possible to create an ultra-mobile device that retains a spatial interaction metaphor that previously required a concrete mechanism for providing visual feedback.*

1.1 MOTIVATION AND RESEARCH APPROACH

During the course of this research, my collaborators and I took a somewhat unconventional approach. Instead of performing a series of design iterations to solve some real-world problem, we instead focused exclusively on creating an ultra-mobile yet still spatial interface.

We started with the idea that mobile devices are not so mobile anymore; recent top-of-the-line mobile phones are considerably larger than the previous generation. For instance, the iPhone is larger than the previous generation of phones and sizes have since increased even more. We set out to imagine a world where this was not the case; where mobile devices continued, as they evolved, to get smaller and smaller, that is, more and more mobile. However, for that to happen something had to go: the touchscreen.

But, we found it was not enough to simply remove the screen. We also wanted to retain the same interaction style as current mobile devices, so users could transfer the knowledge they have from existing interfaces.

Therefore, the challenge became: to design a mobile device that could be arbitrarily small yet retain the spatial interaction style common to touchscreen devices. Since the interface had to exist somewhere, we considered: what if the interface existed only in the user's imagination. We christened such an interface as an *imaginary interface* and used this as the guiding principle to drive our further research.

Once establishing the "game rules" for Imaginary Interfaces, we set out to explore this new interaction style on four fronts. First, what was the low-level (pointing, drawing, etc.) interaction performance that users could achieve? Second, how could such a system be built? Third, how could users learn where targets were located within an imaginary interface? And fourth, what human perceptual abilities were exploited during interaction with an imaginary interface?

1.2 PRINCIPLES OF IMAGINARY INTERFACES

The guiding principle of Imaginary Interfaces is that the interface exists only in the user's imagination. However, this principle can be broken down into two more specific principles. The first creates a link back to touchscreen devices and the second is a function of mobility.

Principle #1 – There is one global spatial layout of interface elements (i.e., each location in space contains only one UI element at a time).

This principle ensures that the underlying layout of an imaginary interface is spatial and not iconic or gestural. Like the touchscreen devices that Imaginary Interfaces derive from, each UI element exists at a location in space. This allows the user to use their muscle memory and visual understanding of the interface to quickly target a location. We choose to use a simple flat 2D interaction plane to correspond as closely as possible with existing interfaces but the concept of Imaginary Interfaces could apply to curved planes or full 3D interfaces if the user could acquire a strong understanding of the spatial layout.

Principle #2 – Interface elements are not visible when you interact with them (i.e., a user must always imagine where the element is).

By not providing a dedicated surface for input and output, but instead relying on the user to maintain an internal "image" of the interfaces, we could shrink the device to the minimum while still retaining the same interaction style. However, because of this we had to limit any mechanism that temporally multiplexed single locations with multiple functions (for instance, long scrolling lists of options) to maintain a level of consistency, and therefore predictability, in the system.

1.3 ORGANIZATION AND SUMMARY OF FINDINGS

Chapter 1 introduces the concept of Imaginary Interfaces and summarizes the contribution of this dissertation.

Chapter 2 surveys existing research related to Imaginary Interfaces and sets the foundation for this dissertation.

Chapter 3 details three exploratory experiments that begin to flesh out what is possible with an imaginary interface and where they might be useful. The first two studies show that participants were able to reproduce simple drawings in mid-air using an imaginary interface and that targeting performance improved when the participant used their non-dominant hand to set the reference frame of the interaction. The final study in this chapter presents initial results into the selection accuracy of a mid-air imaginary interface.

Chapter 4 presents the design of two interaction styles, mid-air and palm-based, each accompanied by a hardware prototype that showcases the interaction style and suggests how a deployable device could be built. The chapter concludes with a controlled lab study comparing the selection accuracy of both interaction styles.

Chapter 5 presents a method for learning a palm-based imaginary interface based on transferring interface knowledge from an existing device to an imaginary interface. This chapter contains two user studies. The first investigates the amount of spatial knowledge users inadvertently acquire while using a touchscreen mobile phone and if that knowledge can be transferred to another interface. The second study comprehensively tests the selection accuracy on the palm compared to a conventional touchscreen mobile device.

Chapter 6 goes beyond the design of Imaginary Interfaces and investigates what human sensing capabilities enable interaction with a palm-based imaginary interface. The chapter contains three formal lab studies that show how performance with a palm-based imaginary interface is affected primarily by visual cues (i.e., watching yourself interact) and when that is not available, by the tactile cues available on the interaction surface, especially when the interaction surface can feel itself being touched, as is the case with a palm-based imaginary interface.

Finally, Chapter 7 wraps up the contributions of this dissertation and presents limitations, several directions for continued work and a discussion on how designers of future mobile devices could apply the lessons learned in this dissertations to create more humane devices that aim only at augmenting our experiences, not replacing them.

2 RELATED WORK

The goal of Imaginary Interfaces is to create mobile interactive interfaces that offered a spatial interaction style and could be fully miniaturized. As such we were heavily inspired by the research on spatial interfaces, which firmly establishes the benefits of this interaction style, and from wearable computing, from which we take the goal of ultra-mobility and convenience.

The wearable computing body of literature is rich in prototype systems that offer extreme mobility by using the body as an interaction surface. Similarly inspiring were mid-air gestural interfaces that offer interaction that does not require a physical device.

Finally, since Imaginary Interfaces have no visual feedback, eyes-free interfaces and interfaces for visually impaired users are especially relevant, as is the large body of perceptual psychology literature on tactile and multi-sensory perception.

2.1 SPATIAL INTERACTION

Modern UIs often take advantage of the human ability to use and learn an interface with spatially arranged UI elements. Especially when compared to lists or hierarchical arrangements of commands, spatially arranged interfaces often perform very well. For instance, in the classic Data Mountain study, Robertson et al. (1998) found that when users could choose the spatial layout of webpage shortcuts on a virtual 3D interface, they were able to retrieve these shortcuts with much higher accuracy than with a list-based interface, both immediately after (Robertson et al. 1998), and in followup sessions (Czerwinski et al. 1999). A subsequent study by Cockburn and McKenzie (2002) showed that the 3D nature of the original interface was actually a hinderance and a simple 2D interface outperformed the 3D version in both virtual and physical interfaces.

For users to properly operate a spatially arranged interface, it needs to be anchored to something. With a conventional touchscreen the frame of the device unambiguously provides this function but for less fixed interfaces, the method of stabilization becomes a design choice.

Billinghurst et al. (1998) categorizes augmented reality (AR) based spatial interfaces as either head, body or world-stabilized. Imaginary Interfaces are, in one sense, body-stabilized as the camera is mounted to the user's chest and tracks the location of both hands with respect to the body. However, we take an additional cue from bimanual interaction research, in particular Guiard's (1987) kinematic chain model, and use the non-dominant hand to coarsely set the frame of interaction, while the dominant hand performs fine input within that space. This effectively creates a fourth style of stabilization (beyond Billinghurst et al.'s three): a non-dominant hand stabilized interface.

Using the non-dominant hand to set the reference frame is well known to help mid-air interaction. Hinckley et al. (1997) found that two hands interacting together provide enough context to maintain a frame of reference without visual feedback. Balakrishnan and Hinckley's (1999) followup showed that either visual feedback from the interface or from the non-dominant hand is sufficient to maintain the *kinesthetic reference frame* necessary for spatial interfaces. Tan et al. (2002) continued this work by showing the kinesthetic cues available from pointing to a target on a touchscreen help to improve performance over mouse-based interaction.

Concurrently to our initial investigation into Imaginary Interfaces, Cockburn et al. (2011) were investigating what they call *air pointing*. They looked at three styles of pointing with varying levels of visual feedback: ray casting onto a 2D plane, directly on a 2D plane and in 3D. The mid-air Imaginary Interface corresponds very closely to the 2D plane air pointing with no visual feedback. Cockburn et al. presented a framework for air pointing designs in two parts. The first part, interaction qualities, specified the goals of an air pointing interface. The goals were: learnablity by novices, selection speed for experts, expressivity (the number of selectable locations), cognitive effort and comfort. The second part identified the interaction dimensions as the reference frame for spatial input, scale of spatial input control, input degrees of freedom, feedback modality and feedback content.

Cockburn et al. did an excellent job of predicting some of the issues we would find with Imaginary Interfaces. For instance, their accuracy measurements for feedback-less interaction are, while worse than with full feedback, at an acceptable level for some interactions. They found the 2D interaction plane to be the most efficient interaction style, which we instinctively chose as the interaction layout for the first version of Imaginary Interfaces. This finding clearly spelled out the need for a strong reference frame, which is one of the main themes of Chapter 3, where we show that using the left hand as the origin to the interaction provides a very strong reference system that can be maintained while in motion.

2.1.1 Extending spatial input beyond the display

Recent research projects in mobile computing have led to a number of input techniques that extend the area for spatial input to the space around a conventional touchscreen. For example, SideSight (Butler et al. 2008), Abracadabra (Harrison and Hudson 2009) and HoverFlow (Kratz and Rohs 2009) (who introduce the term *around the display interaction*) all provide a level of spatial input around a mobile device with the goal of increasing the interaction space. With Piles-across-space, Wang et al. (2009) allow users to create bins of off-screen objects around a handheld device. Jones et al. (2012) explores how to use the space around the device to support multi-scale navigation.

However, this increase in input area can come at a cost. Ens et al. (2011) measured user performance of off-screen selection when guided by on-screen feedback. They revealed that directly pointing to objects off-screen takes up to fours time longer than pointing to a comparable object shown visibly (in this case with a projector) and that pointing accuracy decreases with distance into offscreen space. However, it might be possible to overcome these limitations with more advanced techniques, such as using *spatial correspondence* (matching off-screen and on-screen interfaces) to improve the selection accuracy of unseen targets (Pietroszek and Lank 2012) or using techniques like Around-Device Binning (Hasan et al. 2013) to collect off-screen space into "bins" that improve retrieval time.

Regardless of performance measures, increasing the interaction space beyond the screen offers a compelling interaction style that researchers continue to experiment with. Peephole displays (Yee 2003) and the original concept from Cameleon (Fitzmaurice 1993) offer spatial input that maps physical movement of the device to movement in a virtual world. One can revisit a virtual location by returning to the same physical position. Similarly, Cao and Balakrishnan (2006) use tracked handheld projectors to create a room-sized interface which is only ever partially revealed by the beam of the projector and with Minput (Harrison and Hudson 2010), users move around a small device to operate within a larger space using only the small screen of the device for visual feedback.

Some research projects have explored allowing multiple users to access the interaction space together, which enables some compelling interactions. Cao et al. (2007) extend their handheld projector interaction and Lucero et al. (2010) extend the peephole metaphor to multiple users who combine to create a larger interface, with each mobile device displaying a spatially consistent portion of the global interface. Second Surface (Kasahara et al. 2012) is a collaborative system that allows users to interact in the empty space of the environment around them, while the results of this interaction (for example, 3D drawings) are visible to other users of the system through the device.

2.1.2 Spatial interaction without a screen

Mobile phones no longer follow a trend toward miniaturization and are actually larger today than they were a few years ago. This trend reversal is partly because small touchscreens are very difficult to operate. Projects such as nanotouch (Baudisch and Chu 2009) and Ridgepad (Holz and Baudisch 2010) have showed how to increase touch accuracy on very small screens but the real miniaturization begins when the screen can be eliminated completely. This is the approach we took with Imaginary Interfaces as removing the screen all together allowed Imaginary Interfaces to become arbitrarily small.

This approach has also been followed by a few other projects but for different reasons. The most similar to Imaginary Interfaces is Virtual Shelves (Li et al. 2009; Li et al. 2010), a mobile phone interaction technique that allows users to invoke a command in an eyes-free manner by pointing to it in the hemisphere in front of them. Folmer and Morelli (2012) combined free hand movement with vibrotactile feedback to specify the location of space of an interface element without visual feedback. They called this a tactile-proprioceptive display where users scan the interaction area to come across targets that are signaled with vibrotactile feedback. Similar principles can be applied to mobile phone rotations (Morelli and Folmer 2012) and movements through an environment (Fallah et al. 2012).

Other proposed interfaces offer non-visual spatial interfaces for several specialized scenarios. For instance, Mouseless (Mistry and Maes 2010) allows users to mimic the use of a mouse beside a laptop, Spatial Sketch (Willis et al. 2010) allows users to define the shape of a laser printed object by gestures, Data Miming (Holz and Wilson 2011) captures the description of an object from the user's hand gestures to select objects from a large database. Also, Imaginary Reality Gaming (Baudisch et al. 2013) uses similar principles to enable playing a ball game without the game by obtaining all the information needed from how other players interact with an imaginary ball.

For the most part all non-visual interfaces sacrifice some interactive ability to offer an non-visual interface. They do this to increase mobility by allowing for a smaller, lighter or cheaper device; or when offering an interface for visually impaired users. However, for some interfaces, visual feedback has been found to actually harm interaction performance. Witt et al. (2008) found that visual feedback in a head-mounted display prevent users from attending to the gestures they are performing, which compromised performance. Similarly, Kajastila and Lokki (2013) provides some hint that vision-based interaction might sometimes be a hinderance. Their results indicate that audio-based menus can be comparable to visual menus and are especially useful in situations where the user's vision is occupied with some other task, such as walking or driving.

2.2 INTERACTION ON AND AROUND THE BODY

Imaginary Interfaces derive in spirit from the wide area of wearable computing. We share the goal of creating always available, nearly invisible and completely integrated personal computing technology to support day to day life.

2.2.1 Wearable devices

By integrating a device into users' clothing or attaching it to their body in some way, wearable devices offer a robust method of implementing the goals of wearable computing. For instance, the Body-Coupled FingeRing (Fukumoto and Tonomura 1997) uses finger-worm sensors to turn any surface into an input device and GesturePad (Rekimoto 2001) turns part of the user's clothing into a touch pad.

With Disappearing Mobile, Ni and Baudisch (2009) explored extremely small wearable input devices. They tested text input using a single point scanning interface mounted on the user's wrist. They point out that one of the main challenges for spatial interaction without visual feedback is to connect strokes (*relative position features*). Sturm et al. (2009) obtained a similar finding when investigating gestural interfaces for blind users and use this to motivate for dynamic haptic displays to provide the necessary feedback to support the interaction.

Inspired by the wrist watch, many wearable computing projects have chosen the wrist as a location to mount an interface. Rekimoto's (2001) GestureWrist infers hand gestures indirectly by sensing changes in the shape of the wrist, while Raghunath and Narayanaswami (2002), Blaskó and Feiner (2004), Ashbrook, Lyons, et al. (2008) and among others investigated direct touch a watch-like device. With their nanotouch device attached to the wrist, Baudisch and Chu (2009) explained how touch input precision could be improved by touching the underside of the watch band.

Other systems, such as GestureWatch (Kim et al. 2007) and subsequent follow-up work (Lee and Starner 2009; Deen et al. 2010) as well as Abracadabra (Harrison and Hudson 2009) sense gestures performed by the dominant hand near and around the wrist-worn device. Fukumoto and Tonomura's (1999) Whisper is a wrist-worn speech platform that sends audio signals through the user's finger placed in their ear. Blaskó et al. (2005) built a simulated wrist-mounted projector. Finally, The AugmentedForearm (Olberding et al. 2013) is an exploration into body-mounted displays that extend along the user's forearm and Holz et al. (2012) explore the interaction capabilities of devices implanted under the user's skin.

2.2.2 Interaction on the body

However, the ultimate mobile interface might not be a device at all. The problem with conventional mobile devices, according to Kristoffersen and Ljungberg (1999), is that they often require two hands for input, there is no place to put the device and too much visual attention required. Because of this, one-hand interaction is well-suited for mobile interaction because it leaves the other hand free to manipulate the environment. Kristoffersen and Ljungberg (1999) explain how mobile workers often struggle to operate a handheld device while they are doing their job. Often they must "make place" for their interaction, interrupting the flow of their primary task. One solution to this is to use a readily available surface within the environment as the interaction surface.

The availability of the user's own body as a surface for mobile interaction coupled with a body-worn projectors has been exploited in many research projects. The early concept MARISIL (Pulli and Antoniac 2004) led the way to the more advanced prototype systems Brainy Hand (Tamaki et al. 2009), Skinput (Harrison et al. 2010), Palm Display (Kim et al. 2010), OmniTouch (Harrison et al. 2011) and Armura (Harrison et al. 2012).

Sixth Sense (Mistry et al. 2009), a projector and camera-based wearable computer, is a well-known example of this type of system. It provides spatial gesture input like Imaginary Interfaces but relies heavily on the projector to provide visual feedback.

In situations that do not afford projection, the user's familiarity with their own body allows for non-visual interfaces that exploit the user's tactile and proprioceptive senses. For instance, with their Body Mnemonics design concept, Ängeslevä et al. (2003) provide an earlier example of this interaction. While, BodySpace (Strachan et al. 2007) and Mnemonical Body Shortcuts (Guerreiro et al. 2008) provide prototype systems that use positions on the body as shortcuts for mobile device functionality.

Similarly Point-Upon-Body (Lin et al. 2011) and Shoemaker et al.'s (2010) body-centric wall interaction prototype assign functions to positions on the user's body that can be activated with the help of feedback from an external display (e.g., with interactive walls).

Wagner et al. (2013) goes beyond the system-related work to present a design framework for taking into account how interaction on the body influences interaction in the environment. The design space, called BodyScape, allows researchers to classify techniques based on the relationship of the input and output surface of the environment. For instance, a technique that combines mid-air pointing with on-body touch will conflict when the on-body target is located on the pointing arm.

2.2.3 Interaction on the palm

In the same way, other projects have used the user's palm as an interaction surface but in many cases coupled with a projector for output. Several projects, such as the already mentioned Sixth Sense (Mistry et al. 2009) and Brainy Hand (Tamaki et al. 2009), propose using the palm as an interaction surface. In fact, the use of the palm as a mnemonic device has existed since at least medieval times with the Guidonian Hand (Wikipedia 2013), a system for learning music by placing sequences of notes on parts of the palm.

The palm's abundant tactile features and natural divisions make it useful for representing various grid-like interfaces: as a number pad (Goldstein and Chincholle 1999), television remote control (PalmRC (Dezfuli et al. 2012)), for text entry (KITTY from Kuester et al. (2005)) and for elaborate input/output such as with the Mobile Lorm Glove (Gollner et al. 2012). Nakatsuma et al. (2011) alternatively chose to use the back of the user's hand as the interaction surface.

In virtual reality (VR) applications, the palm plays a central role in some techniques. For instance, the Haptic Hand (Kohli and Whitton 2005) exploits the ready availability of the user's non-dominant hand to provide a surface for interaction in VR applications. The performance improvement possible when interacting on a surface, instead of in the mid-air was also the motivation behind Sibert and Hahn's (1999) Hand-Held Windows, which used a physical prop as an interaction surface while interacting in a VR environment. This phenomenon was studied by Wang and MacKenzie (2000), who found that having a physical constraint (in their case a table) improved VR interaction speed substantially with a small decrease in accuracy.

2.2.4 Interaction around the body

Other interface concepts have exploited users' intimate familiarity with their peripersonal space (defined by Rizzolatti et al. (1997) as the space directly around us) and their proprioceptive abilities. Chen et al.'s (2012) collection of body-centric interaction techniques shows how the space on and around the body can be combined to offer compelling interactions. Folmer and Morelli's (2012) previously mentioned proprioceptive display combines proprioception with spatially triggered vibrotactile feedback to allow eyes-free exploration of the featureless space in front of the user.

Similarly, Motion Marking Menus (Oakley and Park 2009) use proprioception to enable eyes-free input for handheld devices and the AirWriting (Amma et al. 2013) glove uses an accelerometer and gyroscope to effectively capture and characterize in-air writing for text entry. While the Wearable Virtual Tablet (Ukita and Kidode 2004) enabled users to virtually draw on arbitrary surfaces using a head mounted camera.

2.2.5 Microinteractions

Imaginary Interfaces (and especially the Imaginary Phone introduced in Chapter 5) are particularly well suited for supporting *microinteractions*—the quick mobile device interactions that characterize the dominant interaction mode for mobile phones (Ashbrook 2010).

Ashbrook, Clawson, et al. (2008) showed that it takes over 4.5 seconds on average just to begin an interaction with a mobile phone stored in your pocket. This is a substantial overhead for an interaction that overall often only lasts a few seconds. Oulasvirta et al. (2005) showed that in many situations mobile devices users only have 4–8 seconds to complete an interaction before they must attend to the environment around them. To satisfy this constraint, Ashbrook (2010) argues for systems based on sequences of microinteractions that each take less than four seconds from start to finish.

Several research projects have explored interaction techniques that were either designed specifically to support microinteractions or could be applied to them. For instance, with PinchWatch, Loclair et al. (2010) proposes using a depth camera to sense one-handed gestures, coupled with a small wrist-mounted display, to support microinteractions. Whack gestures (Hudson et al. 2010) are also well-suited for microinteractions. They are super quick interactions performed by slapping a device while it remains in your pocket—for example, to silence a ringer. PocketTouch (Saponas et al. 2011) also offers quick interaction by sensing touch on the surface of a device through clothing. Nenya (Ashbrook et al. 2011) employs a magnetically tracked ring to sense small movements of the ring that could be used as a low-bandwidth input device suitable for microinteraction-style tasks.

Wolf et al. (2011) explored *microgestures* to perform microinteractions while grasping another objects and with Tickle (Wolf et al. 2013), proposed using finger-mounted sensors to detect these small gestures while interacting with any object.

Lower fidelity gestures, such as head tilting (Crossan et al. 2009) or wrist movement (Rahman et al. 2009) are also good candidates for microinteractions.

2.3 SENSING TECHNIQUES FOR INTERACTION

There are many approaches to sensing interaction on and around the user's body that would be applicable to Imaginary Interfaces. Approaches typically use either custom electronic hardware to sense the interaction or employ computer vision tehcniques to interpret the scene in front of the user including the user's hand gestures. See also Morris et al.'s (2010) survey on *always-available* mobile interaction as an excellent starting point.

2.3.1 Hardware-based approaches

There are a few technical approaches to sensing touch on the surface of the hand. First, a designer could place sensing material on the palm. For example, the Chording Glove (Rosenberg and Slater 1999) and KITTY (Kuester et al. 2005) cover parts of the operator's hand with electrical contacts that, when touched with another contact, register a touch event at a specific location. The Mobile Lorm Glove (Gollner et al. 2012) similarly uses fabric pressure sensors distributed over the surface of a glove to sense touch. Although this method could produce highly reliable and perhaps high-resolution input, the fact that the user must wear something over their hand prohibits the general use of this approach.

Second, a system could observe the physical manifestations of touch from afar. For example, Skinput (Harrison et al. 2010) senses taps on the hand and forearm by measuring the different patterns of vibrations that travel up the arm. Point-Upon-Body (Lin et al. 2011) uses ultrasound transducers located on the users' wrists to localize touch on their forearms. Nakatsuma et al. (2011) show how to sense touch and gestures on the back of the hand with infrared (IR) proximity sensors and a piezoelectric transducer. Touché (Sato et al. 2012) uses swept frequency capacitive sensing obtain a unique capacitive attenuation profile that differs based on the path current takes through the body. They use this to train a machine learning classifier that classifies touch on various locations on the body. The Magic Finger (Yang et al. 2012) uses a small camera attached to the end of the user's finger to enable interaction on the body or any other surface.

Similar techniques can also be used to sense the hand posture and gestures. GestureWrist (Rekimoto 2001), for example, senses hand postures by observing the changes in the shape of the wrist, while Hrabia et al. (2013) reconstructs hand pose from sensors located on the user's finger joints.

2.3.2 Computer vision-based sensing

Alternatively, computer vision has been used for a wide range of hand sensing interfaces. However, each computer vision sensing technique tends to cater to specific situation. There is a large and vibrant research community built around the problem sensing articulation of a user's hands with computer vision; the survey on this topic from Moeslund et al. (2006) contains 424 references and research has continued strongly since then. However, there is currently no general purpose hand tracker that recognizes the pose and location of hands in a natural environment. Such a solution might not be too far off, as recent research has demonstrated some very impressive results. For instance, Oikonomidis et al. (2012) present a system that successfully separates two intertwined hands, a difficult problem because of the inherent substantial occlusion.

From a more application centric point of view, Wachs et al. (2011) present a survey of vision-based approaches to hand gesture application. They argue that computer vision based approaches are well suited for hand gesture interaction because the sensing is inherently nonintrusive, passive and silent, and available at a low cost. Furthermore, the installed camera can be used for other tasks (such as taking photographs) beyond sensing interaction. However, there are significant challenges to the wide-spread adoption of vision-based sensing: robust sensing is very difficult to obtain in varying lighting condition, moving scenes and varying background and because of differences among users.

Researchers have explored many on-body locations and configurations for a wearable camera, from the chest-mounted form factor we use in our Imaginary Interfaces prototypes, to head-mounted (Tamaki et al. 2009) and shoulder-mounted (Harrison et al. 2011) and even shoe-mounted cameras (Bailly et al. 2012). Mayol-Cuevas et al. (2009) present a framework for evaluating the placement of wearable vision sensors on the user's body. They argue that, although there is no overall best placement, an active camera (i.e., mounted on a servo to allow it to follow a target of interest) mounted on the shoulder is an excellent compromise between field of view, social acceptance, suitability for a range of tasks and body motion.

Digits (Kim et al. 2012), and predecessors (Vardy et al. 1999; Ahmad and Musilek 2006), takes a slightly different approach by using a wrist-mounted camera to determine the posture of the hand, but not its position in space, to support gestural input. While Wang and Popović (2009) demonstrated robust hand tracking with computer vision based on a multicolored glove.

Until recently, computer vision based sensing was limited to tracking the location and pose of each hand for the purpose of gestural interaction, but they did not allow the user to use their uninstrumented finger to select a position on their bare hand. This, perhaps, has not been done previously because of the difficulties of separating the two hands. With the wide-spread available of depth sensing cameras this has become tractable. Wilson (2010) showed how to detect touch on static surfaces using a depth sensing camera, while OmniTouch (Harrison et al. 2011) and the Imaginary Phone (from this dissertation) use computer vision techniques with a wearable depth sensing camera to interact with an interface located on the users hand and forearm.

2.4 gestural and 'natural' interaction

As Jacob et al. (2008) explain, interactive systems have evolved substantially over the years. Starting with command-line interfaces, moving to direct manipulation (Shneiderman 1983) and instrumental interaction (Beaudouin-Lafon 2000), we are currently entering the era of Reality-based Interaction. Jacob et al. described how Reality-based Interaction is a new model for interactive devices and techniques that exhibit four complementary components: 1) Naïve physics; 2) Body awareness and skills; 3) Environment awareness and skills; 4) Social awareness and skills. These concepts are similar to the term natural user interface (NUI), which is often used to note the fundamental shift from metaphor-laden graphical user interfaces (GUIs) to a more direct and intuitive interface.

Imaginary Interfaces sits between categories. It employs the free hand gesture style common in NUIs but by retaining spatial interaction, it allows users to leverage their familiarity with direct manipulation touch interfaces.

By allowing for empty handed interaction, Imaginary Interfaces takes advantage of the naturalness of gestural interaction. Gesture is complex behavior that has been widely studied in other disciples (see McNeill (1992) for example) but for the purposes of interaction with a computer system, as suggested by Mulder (1996) we can use Sturman and Zeltzer's (1993) definition of whole hand input as "the full or direct use of the hand's capabilities for the control of computer-mediated tasks". I adopt a slightly more expanded definition to also include the use of the whole body.

This interaction style is not all that new. Starting in the 1980s with Bolt's (1980) 'put-that-there' and Krueger et al.'s (1985) VIDEOPLACE researchers have been investigating ways of increasing the expressiveness of interaction by using gestures performed by users' hands and their bodies as input to computer systems. See also the gesture taxonomy for interaction from Karam and Schraefel (2005).

However, only recently, motion sensing components (such as accelerometers and gyroscopes) have became routinely embedded in mobile and wearable devices, opening up this input channel to control our devices. At the same time, advances in depth sensing cameras (most notably with Microsoft Kinect's full body skeleton reconstruction (Shotton et al. 2011)) have made full-body gestural input common place, at least for home entertainment.

Also, for years, immersive VR environments have used freehand gesture input (i.e., whole hand interaction (Sturman and Zeltzer 1993)) as their primary interaction mode. Pinch-sensing gloves, such as the DataGloves (Zimmerman et al. 1987), have been used since the 1980s, see Sturman and Zeltzer (1994) for a detailed survey. Even ignoring the cost and complexity, gloves are limited in their general usefulness because of social norms and problems with durability.

Researchers have also transferred this interaction style to the desktop. GWindows (Wilson and Oliver 2003) is a desktop gestural interface based on robust stereo vision techniques. And projects such as Charade (Baudel and Beaudouin-Lafon 1993) have investigated how humans can interact within virtual environments as they do in person. Others have looked at improving 3D gestural interaction by tracing the gestures over physical props (Jackson and Keefe 2011).

Often though, gestural input is limited to a series of categorical gestures. That is, one posture of the hand results in one corresponding action. The Gesture Pendant (Starner et al. 2000) and predessors (Starner et al. 1997) are classic examples of this interaction style. However, these simple gestures do not approach the complexity and expressiveness of gestures in normal social interaction, which support a conversation by adding a secondary stream of information that complements speech with spatial relationships that would be difficult to express otherwise (Emmorey et al. 2000).

2.5 MOBILE INTERFACES FOR VISUALLY IMPAIRED USERS

Many systems have been developed to help visually impaired users operate the predominately visual interfaces present on modern computing devices. Since these systems all operate without visual feedback, they are particularly relevant to Imaginary Interfaces.

Visually impaired users rely heavily on tactile cues but modern touchscreen-based devices lack the tactile discoverability of button-based devices. To address this, McGookin et al. (2008) investigated tactile overlays and gesture-based interfaces to increase the usability of touchscreen phones. The Talking Tactile Tablet (Landua and Wells 2003) uses tactile and audio feedback to complementarily reinforce learning through dual modalities. EarPod (Zhao et al. 2007) and BlindSight (Li et al. 2008) combine liberal amounts of audio feedback with a tactile-rich form factor to enable eyes-free operation.

Touchscreen-based interfaces allow for highly dynamic interfaces where the user cannot predict where a given function will be located. To address this, researchers have turned to audio feedback to "explain" the interface to the user. For instance, Pirhonen et al. (2002) investigated combining audio output with gestural input, and Brewster et al. (2003) followed up on the work by improving the audio feedback with 3D spatiality and a more dynamic nature.

Text entry provides an even more difficult problem with visually impaired users, especially on touchscreen-based devices. Oliveira et al. (2011) compares several recently proposed touch-based text entry methods. They found that performance with a given technique depends heavily on the user's abilities (tactile sensitivity and spatial reasoning) and there is no overall best technique.

Beyond research prototypes, visually impaired users regularly employ mobile technology to gain more independence (Kane et al. 2009). Commercially available mobile phone interfaces come in two categories: *cursor-based* and *touch-and-explore* interfaces.

Cursor-based interfaces, such as Mobile Speak[1], have a cursor that announces the current function as the user moves around the interface in single steps, allowing the user to traverse the interface in a predictably way.

Alternatively, touch-and-explore interfaces allow users to navigate the interface by dragging freely on the touch screen and listening to the auditory feedback in response (as in the Talking Fingertip Technique (Vanderheiden 1996), SlideRule (Kane et al. 2008), VoiceOver for iPhone[2] and Explore by Touch for Android devices[3]). The touch-and-explore interaction mode allows users to access familiar items faster than the linear effort imposed by a cursor-based list. However, to do this, they must build up spatial memory to be able to target a memorized location.

1 http://www.codefactory.es/en/products.asp?id=316
2 http://www.apple.com/accessibility/iphone/vision.html
3 http://support.google.com/android/bin/answer.py?hl=en&answer=2492750

2.6 PSYCHOLOGICAL FOUNDATIONS OF NON-VISUAL INTERFACES

Since Imaginary Interfaces rely so much on unconventional feedback mechanisms (tactile, etc.) it is important to survey the wealth of knowledge in this area available from psychology literature.

2.6.1 Transfer learning

The Imaginary Phone, introduced in Chapter 5, depends heavily on transferring spatial knowledge between interfaces. However, this is not the only place where transfer learning has been applied. In education, transfer of learning is the concept of learning in one context either enhancing or undermining performance in some other context (Perkins and Salomon 1992) and we know that knowledge is best transferred when the circumstances of acquisition are similar to retrieval (Morris et al. 1977).

In computing, transfering spatial knowledge from one context to another is often used. New versions of software depend on being similar enough to previous versions, so users can transfer there knowledge and remain productive. Also, researchers have looked into how well knowledge transfers between environments. For instance, Wallet et al. (2009) investigated the transfer of spatial knowledge from a virtual environment to the real world. Such a system could be used for training of dangerous or impractical situations. Hornof et al. (2008) looked at transferring spatial knowledge from a 2D visual virtual environment to a 3D auditory space.

2.6.2 Spatial memory and visual perception

Baddeley and Hitch's (1974) model of working memory contains a subsystem called the *visuospatial sketchpad* that is responsible for maintaining short term visual memories and spatial relationships. Imaginary Interfaces takes advantage of this ability and throughout this dissertation we will refer to people's short term capacity to recall the location of objects in space as *visuospatial memory*.

When pointing to a target in mid-air, users perform best if they can see both the target and their hand (Berkinblit et al. 1995). However, when removing the image of the target (as with our mid-air imaginary interface style) a similar perceptual process is in place, as the remembered target can be placed into the well established current frame of reference (Darling and Miller 1993), leading to only slightly degraded performance.

2.6.3 Multi-sensory integration

When interacting spatially in the world, humans gather information from many senses (visual, tactile, proprioceptive, etc.) that must be combined (using a process called *multisensory integration*) to produce a general understanding of the environment (Driver and Spence 1998). Because of this, even though modern touchscreen interfaces rely heavily on vision, proprioception and taction also play an important role. For details on the mechanisms of multisensory integration refer to the excellent review article by Ernst and Bülthoff (2004).

However, proprioception alone is not precise enough to enable fine-grained interaction (targeting can be off by 80 mm on average (Fuentes and Bastian 2010)). Instead, eyes-free interaction typically involves proprioception and taction working together since taking either away degrades performance substantially (Voisin et al. 2002). Similarly vision and taction work in concert, at least for the hand (Làdavas et al. 2000).

Touch itself is multi-faceted and has three distinct flavors: *active touch* (the person touching something); *passive touch* (something touching the person); and *intra-active touch* (the person touching him or herself) (Bolanowski et al. 2004). Each has its own capabilities: active touch is a scanning mechanism that allows the actor to build up an understanding of the scene over time (Gibson 1962), while passive is limited to "being touched". However, this is mitigated by the high spatial resolution of the hand (tactile discrimination ranges from 7.7 mm on the palm to 1.6 mm on the index finger tip (Vallbo and Johansson 1978)). On the other hand, intra-active touch, as is used in palm-based imaginary interfaces, combines the capabilities of both (Bolanowski et al. 2004), allowing users to actively explore the interface while passively noting the location of discovered targets.

3 IMAGINARY INTERFACES

In this chapter, I begin the exploration of what is possible with Imaginary Interfaces by asking the question: To what extent can users interact spatially with an interface that exists only in their imagination?

I present three studies that begin to get at the essence of interacting with Imaginary Interfaces. The first study investigates participants' ability to create simple drawings in mid-air; the second investigates how well participants could annotate simple drawings when stationary compared to when mobile; and finally, the third study obtains a first understanding of pointing accuracy with this style of interaction.

Contributions

The main contribution of this chapter is to elaborate on the concept of Imaginary Interfaces by providing the results of early proof-of-concept studies on how well users could interact with an interface that exists only in their imagination. I show that, with a simple mid-air imaginary interface, 1) users are able to create simple drawings, in some cases better than with comparable eyes-free interfaces; 2) using the non-dominant hand as a reference frame significantly improves interaction accuracy when mobile; and 3) users are able to consistently select mid-air targets based on coordinate instructions, albeit with relatively high error.

3.1 PROOF OF CONCEPT USER STUDIES

Each of the following studies investigated participants' ability to interact spatially using an imaginary interface. Since the distinct feature of Imaginary Interfaces is that there is no visual feedback, our main goal across these studies was to determine to what extent users' visuospatial memory (i.e., their ability to recall precise spatial locations in their visual field) could replace the visual feedback traditionally present on mobile devices (e.g, on a touchscreen). In other words, our goal was to determine if users could define an imaginary interaction plane in mid-air and successfully manipulate and annotate objects located on that plane by imagining where the objects were located.

The first study investigated if participants could draw single stroke characters and simple sketches. Participants completed the single stroke characters with high (94.5%) recognition rate. However, the quality of multi-segment drawings decreased with the number of strokes, presumably because of the limitation of their visuospatial memory.

The second study investigated how far user motion impacts visuospatial memory. Participants' performance of annotating an imaginary drawing dropped significantly when they physically turned around between drawing and annotating. However, by using their non-dominant hand as a spatial reference, the effect was partially alleviated.

The third and final study tested participants' ability to point to a location specified in Euclidian coordinates, specifically vertical units of index fingers and horizontal units of thumbs. We found targeting error correlated with the Manhattan distance of the target from the user's fingertips, which serve as strong visual landmarks.

In each experiment we used sensing hardware (an optical tracker installation), which is not appropriate for a real life deployment. Using this highly reliable and accurate platform allowed us to test the underlying limits of user performance with the interaction style instead of the quality of a prototype implementation.

3.2 STUDY 1: BASIC SHAPES AND DRAWINGS

To obtain an understanding of the role of visuospatial memory for interacting with imaginary interfaces we set out to discover if users of imaginary interfaces are subject to the same difficulties in connecting strokes reported by Ni and Baudisch (2009). With a similar goal of ultra-miniaturization, they tested text input using a single point scanning interface mounted on the user's wrist. They discovered that one of the main challenges for spatial interaction without visual feedback is to connect strokes, as they lack *relative position features*.

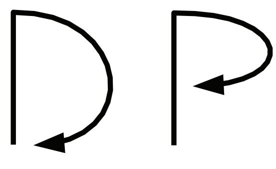


Figure 3.1: In the study reported by Ni and Baudisch (2009), 'D's were often misrecognized as 'P's.

As illustrated by Figure 3.1, their participants often drew what was recognized as a 'P' character instead of a 'D' because the participants failed to connect the end point of the stroke to the starting point due to the absence of visual control. The participants could not see the interaction with the sensor because it was obscured by their own hand.

The current study replicates that experiment using an imaginary interface to determine if the participants' visuospatial memory can help them locate the relative position features and enable drawing of correctly formed shapes.

3.2.1 Tasks and procedure

After an introduction to the task and apparatus, each participant began a series of trials. For each trial, the experimenter held up an A4 sheet of paper to the participant that contained one of the randomly selected stimulus drawings for the current task. The stimulus drawing are reproduced in Figure 3.2.

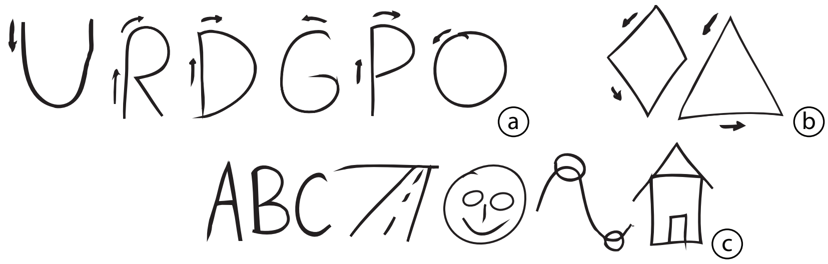


Figure 3.2: Study 1 stimulus drawings – for each task, participants reproduced (a) task 1: single-stroke Graffiti characters; (b) task 2: simple shapes drawn repeatedly; (c) task 3: more complex sketches involving multiple strokes.

When the participant indicated they were ready, the experimenter hid the stimulus drawing from view and the participants replicated it in the space in front of them as illustrated by Figure 3.3a. Participants performed an 'L' gesture with their left hand and sketched with their right hand. The system recorded the 3D position of both of the participant's hands throughout the trial. The trial was complete when the participant indicated so.

The trials were divided into three tasks, performed in order:

1. *Graffiti:* Participants drew six Graffiti[1] characters. The characters, shown in Figure 3.2a, were selected from Ni and Baudisch (2009) because they were especially difficult to complete eyes-free. The challenge was to connect and align the end of the stroke to obtain proper recognition.
2. *Repeated drawing:* In this task participants drew a simple square and triangle (stimulus is shown in Figure 3.2b) repeatedly five times in a row without stopping—each subsequent drawing placed over the previous.

1 http://en.wikipedia.org/wiki/Graffiti_(Palm_OS)

3. *Multi-stroke drawing:* In this task participants drew a set of five more complex sketches (see Figure 3.2c), each of which involved multiple drawing strokes.

Each participant completed all three tasks, i.e., they drew all the sketches shown in Figure 3.2. They completed the experiment session within ten minutes.

3.2.2 Hypotheses

This study was mainly exploratory in nature but we did have one quantitative hypothesis, namely that participants would perform fewer Graffiti recognition errors than reported by Ni and Baudisch (2009). Even though the sketches had no visual representation, we expected participants to build up visuospatial memory of the shape by watching their hands act, which would allow them to successfully complete the shapes by matching the required relative position features.

The purpose of the *repeated drawing* task was to allow us to obtain a measure of the lower limit of error on connecting (the vertices of the shapes) by using strokes with very simple movements and very short time periods. The purpose of the *multi-stroke drawing* task was to explore how stroke connection accuracy decreases with increasing number of strokes.

3.2.3 Apparatus

The apparatus is shown Figure 3.3b. In order to obtain full 3D position and rotation information we used an optical tracking system (an OptiTrack motion capture system with eight V100 cameras). Participants wore a marker set on the back of each of their hands. A third marker set placed on the participant's sternum allowed us rotate the collected 3D position data to a common orientation for all users removing the effect of rotation. The system tracked the marker sets with 1 mm accuracy.

3.2.4 Participants

We recruited 12 participants (five female) from our institution to take part in the study. They were between the ages of 20 and 30 ($M = 24.2$, $SD = 2.95$). All participants set the reference frame with their left hand and drew with their right. One participant was left-handed (participant 1) but was proficient with using a mouse in his right hand and comfortable performing the drawing tasks with his right hand. Participants were given a small gratuity for their time.

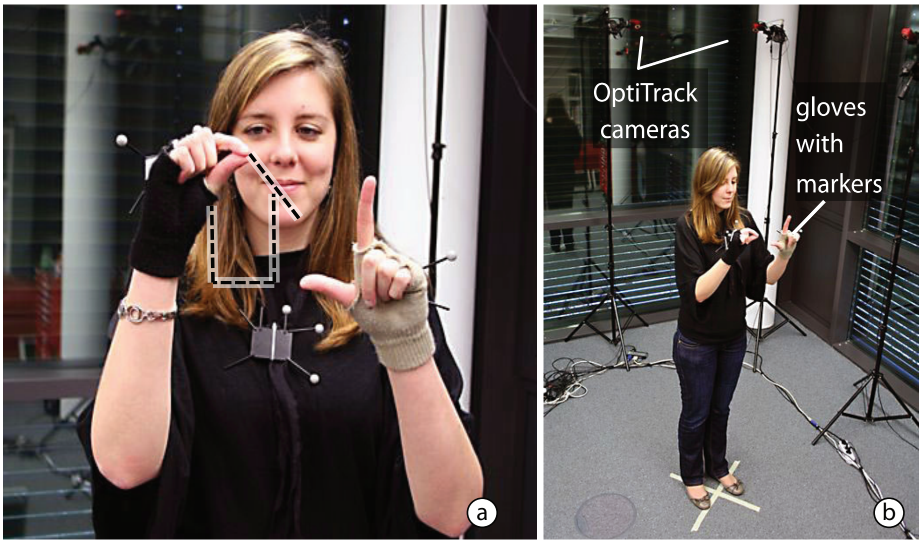


Figure 3.3: Study 1 task – (a) For each trial the participant replicated a simple sketch on an imaginary interaction plane. (b) An optical tracking system tracked hand positions over time.

3.2.5 Data processing

From each trial we collected a series of timestamped 3D positions and rotations for marker sets located on the participant's hands and their sternum. For each timestamp we converted the three 3D positions and rotations to a single 2D coordinate by first rotating them as if the participants were directly facing the *x*-*y* plane, i.e., their gaze followed the *z*-axis. The *z* coordinate was then discarded, leaving the remaining *x* and *y* coordinates to specify the location of both hands on a perfectly vertical interaction plane. The final 2D drawing position was taken relative to the left hand.

The apparatus allowed us to track the position of participants' hands, but not the pinch gestures. Therefore, we manually marked the beginning and end of each stroke post-hoc. Most participants paused briefly at the start and end of each drawing motion, which simplified the classification.

We analyzed the *graffiti* task by running the captured drawings through a Graffiti recognizer[2]. The *repeated drawing* task was analyzed by measuring the average distance per vertex of the triangle/diamond. *Multi-stroke drawings* were not formally analyzed.

2 Pen stroke recognition: `http://blog.monstuff.com/archives/000012.html`

3.2.6 Results

Graffiti task

Figure 3.4 shows the complete set of drawings created by the participants for this task. Overall only 5.5% of the gestures were unsuccessfully recognized versus 15.0% for the same subset from Ni and Baudisch (2009). These results are summarized in Figure 3.5.

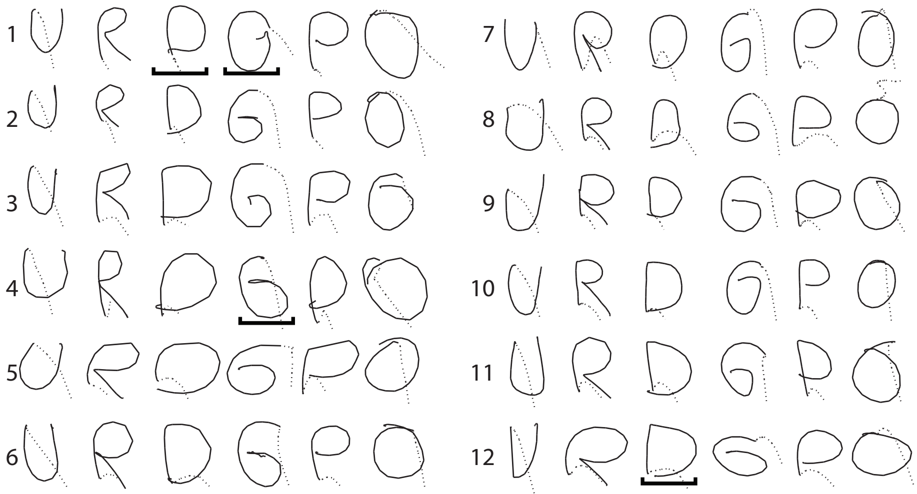


Figure 3.4: Study 1 results for task 1 – all characters participants drew for the *graffiti* task, resized to match. Each half row is data from one participant. Misrecognized characters are underlined.

The error rate of drawing characters using the imaginary interface was comparable to Graffiti text entry on pen devices (2.9% for the same subset of characters by first time users with 5 minutes of training (Mackenzie and Zhang 1997)). While we did not collect enough data for statistical analysis, the data suggest that the *relative position features* issue pointed out by Ni and Baudisch does not apply to Imaginary Interfaces, at least not to the same extent. Since the participants were able to watch themselves interact (and could not in Ni and Baudisch's study), they were likely able to exploit their visuospatial memory to successfully connect the strokes.

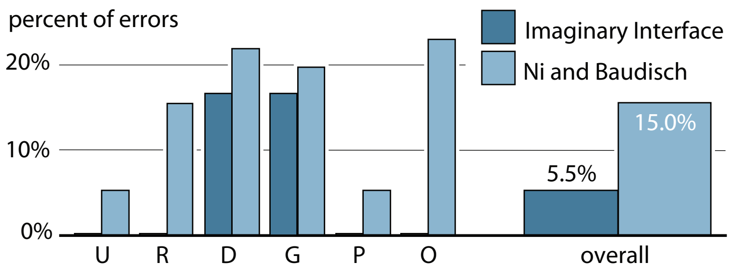


Figure 3.5: Study 1 aggregate results for task 1 – Graffiti recognition error rates compared to Ni and Baudisch's feedback-less gesture input.

The reader must be cautioned that our results presented here and those from Ni and Baudisch were collected under different study conditions. Since we did not test Ni and Baudisch's interface along with our interface, we cannot say conclusively that one is more accurate than the other. However, we think there is good reason to believe that being able to observe the interaction resolves the issue of unmatched position features experienced in Ni and Baudisch's study.

Repeated drawing task

Figure 3.6 shows the complete set of drawings created by the participants for this task and Figure 3.7 shows the average distance between all vertexes for (a) each repetition from the first and (b) each repetition from the previous. Overall the average error from previous for the diamond was 2.20 cm (SD = 0.90) and for the triangle was 3.25 cm (SD = 2.31). The decreasing error distance from the previous suggests that participants built up visuospatial memory of the shape with repetition and the fact that users kept drifting away from the first suggests that later loops overwrote earlier ones.

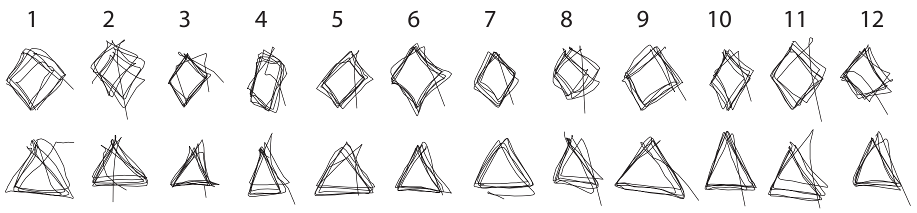


Figure 3.6: Study 2 results for task 2 – all drawings made by all participants. Each column is data from one participant.

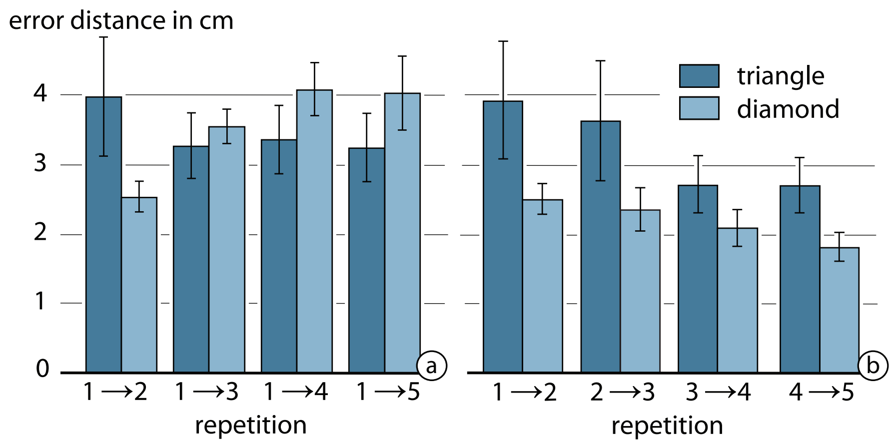


Figure 3.7: Study 1 results for task 2 – graph showing all error rates by repetition (± one standard error of the mean): (a) error rate relative to first drawing and (b) error rate relative to previous drawing.

Multi-stroke drawing task

Figure 3.8 shows all sketches created by the participants. We observed several interesting points: (1) Alignment between strokes seems to decrease with the complexity of the sketch. (2) The individual letters of the ABC string are well drawn; participants, however, condensed whitespace, causing letters to overlap. (3) Relative scale appears reasonably correct. Misalignment appears to be caused mostly by translation errors derived from choosing the wrong starting point.

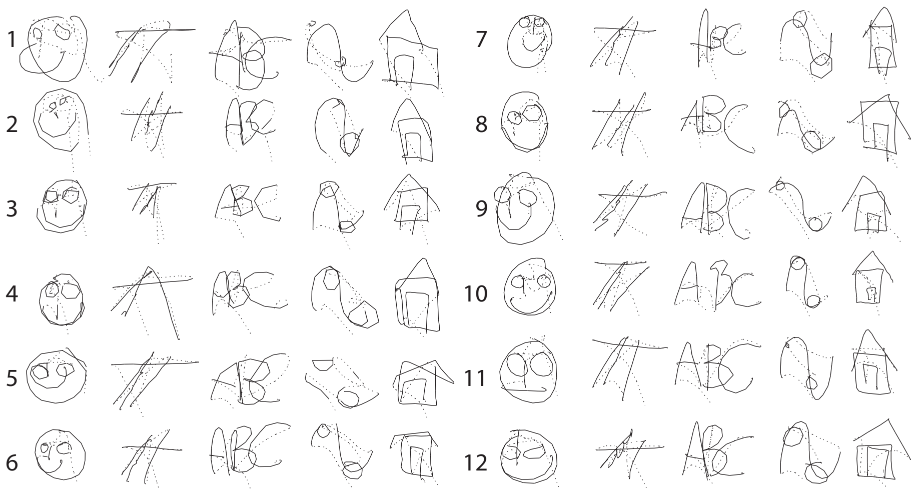


Figure 3.8: Study 1 results for Task 3 – the drawings made by all 12 participants in *multi-stroke drawing* task. Each half row is one participant, resized for comparison.

3.2.7 Discussion

Overall, the results of this initial investigation were promising. Participants were able to create the basic characters and sketches despite the lack of visual feedback. Unlike previous finding, problems with closing shapes were minimal, perhaps because participants built up visuospatial memory by watching their hands throughout the interaction and used this spatial understanding to match the beginning and end of their strokes.

While alignment within a stroke was good, aligning strokes in multi-stroke sketches seemed to challenge participants. There are clearly limits to how complex a drawing can be successfully created without visual feedback. Once a stroke was completed the participants would often slightly misplace the next strokes indicating that their visuospatial memory was beginning to break down. However, it remains unclear under what conditions this breakdown occurs. In the next study, we investigate this more formally.

3.3 STUDY 2: RETURNING TO A DRAWN FEATURE

The purpose of this study was to more formally explore what causes visuospatial memory to fade when using Imaginary Interfaces. In particular we sought to determine if movement caused the participants to misplace their starting point, which would result in misaligned interaction. The participants' task was to draw a simple shape, then go back and point to one of the vertices of what they just drew. In one condition, participants pointed right away; in another condition participants had to turn around between drawing and pointing. This task represents the interaction required when adding to or annotating an existing drawing. We expected that since movement of the participant will change their visual field, their visuospatial memory will also be disrupted, causing their annotations to be worse than when remaining stationary.

3.3.1 Task

Participants began each trial with their hands at their side and the experiment apparatus displayed a simple glyph on a computer screen (Figure 3.9a). As shown in Figure 3.10, all glyphs were constructed from four strokes.

Participants pressed the footswitch to begin the trial and, if in the REFERENCE HAND condition, raised their left hand to perform the 'L' gesture. Then participants replicated the displayed glyph with the right hand (Figure 3.9b) and pressed the footswitch again. In half of the trials, participants rotated their bodies by 90° at this point.

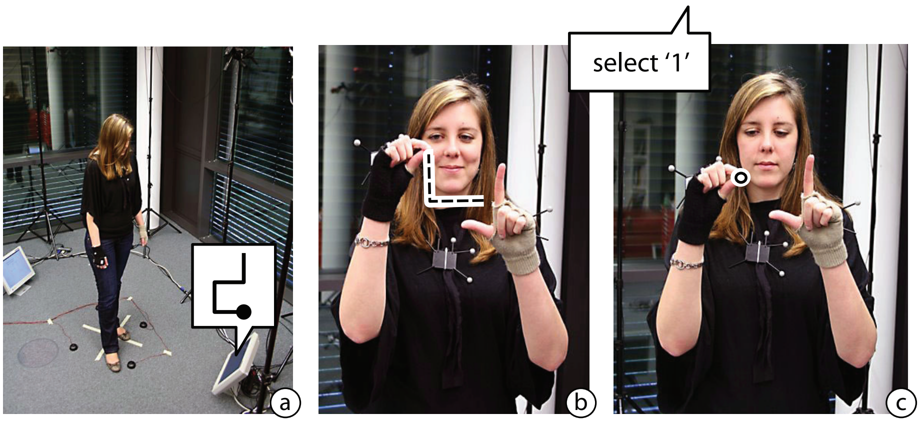


Figure 3.9: Study 2 task – (a) at the start of each trial the participant stood in front of one of the foot switches and viewed the glyph on a monitor. Each trial consisted of two phases: (b) drawing the glyph and (c) selecting a corner of the drawn glyph.

Next a computer generated voice announced "Select 1", "Select 2" or "Select 3" and started the timer. Participants selected the indicated corner from the glyph they just had drawn (Figure 3.10b) and committed by pressing the foot switch, which stopped the timer.

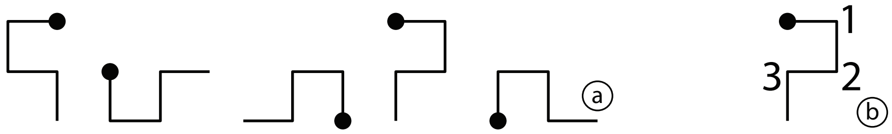


Figure 3.10: During each trial, participants (a) drew one of these five glyphs and then (b) pointed to one of its three corners.

Participants dropped their arms to get ready for the next trial and if they had not rotated earlier in the trial they rotated now.

The system recorded the participant's rendition of the glyph and the selected position. Error was defined as the 2D Euclidean distance between the selected position and the location of the corner drawn earlier. Task time was the duration of corner selection activity.

3.3.2 Independent variables

There were two independent variables: the two-level variable Body rotation and the two-level variable Reference system.

BODY ROTATION In the ROTATE conditions, users rotated their entire body by 90° between drawing the glyph and acquiring the point on it. In the STAY conditions they did not.

REFERENCE SYSTEM In the HAND condition, participants drew the glyph with their reference hand in the air and position was computed relative to this hand. In the NONE condition, they drew with their left hand along their side and position was computed relative to their torso.

3.3.3 Hypotheses

We expected that participants would be able to fully use their visuospatial memory when stationary (i.e., in the STAY condition), but that this would be impaired when moving, i.e., the body rotation done in the ROTATE condition would partially prevent the participant from matching up their current position to their previous position. Therefore we hypothesized that:

HYPOTHESIS H1 There will be lower error in the STAY conditions than in the ROTATE conditions.

However, when rotating, based on Hinckley et al.'s (1997) assertion that using two hands together form a stable reference system, we expected participants will be able to use their left hand as a visual reference point to, in part, fill in for the disruption of the visual reference frame caused by the body rotation.

HYPOTHESIS H2 In the ROTATE conditions, we expect lower error in the HAND condition than in the NONE condition.

Although we had no expectation for a difference in performance between HAND and NONE conditions when the participant did not rotate, we were interested in seeing if an effect was noticeable. We are also curious if the HAND and NONE conditions would differ in task time.

3.3.4 Apparatus

The apparatus for this study was identical to that used in the previous study except that we added two monitors on the floor and four footswitches arranged in a circle around the participant as shown in Figure 3.9a.

We processed the data from the optical tracker in the same manner as in Study 1.

3.3.5 Participants

We recruited a new set of participants, ten males and two females, from our institution and community to participate in this study. They were between 22 and 31 years old ($M = 24.2$, $SD = 3.3$). All were right handed. Each received a small gift in exchange for their time.

3.3.6 Experiment design

The experiment used a two body rotations (ROTATE, STAY) × two reference systems (HAND, NONE) within-subjects factorial design. Each condition consisted of 15 trials. Each trial within a condition used one unique combination of glyph (see Figure 3.10) and corner number. The presentation order within a block was randomized and the blocks were counterbalanced using a balanced Latin square.

With 12 participants, four conditions and 15 trials per condition, a total of 720 trials were completed in the study. Due to tracking errors with the 3D tracking system, 16 of those trials (2.2%) were discarded and not included in the subsequent analysis.

Participants were trained on the operation of the experimental apparatus and the task until they indicated they were comfortable and understood what was required of them. All participants completed the study in 30 minutes or less.

3.3.7 Results

We performed a multivariate 2×2 repeated measures analysis of variance (ANOVA) on error and task completion time. Figure 3.11 contains a summary of the results.

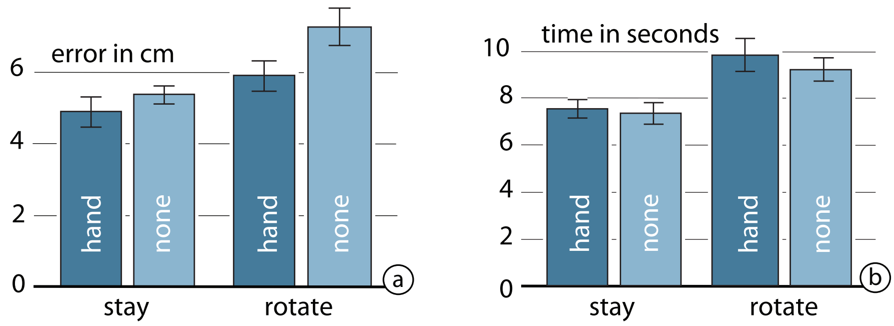


Figure 3.11: Study 2 results – (a) error and (b) task completion time separated by BODY ROTATION and REFERENCE SYSTEM (± one standard error of the mean).

Overall, the HAND conditions had an average error of 5.4 cm (SD = 0.4) and the NONE condition had an average of 6.3 cm (SD = 0.3), which was a significant difference ($F_{1,11} = 16.007$, $p = 0.002$).

Also, participants in the STAY condition (M = 5.1 cm, SD = 0.3) had significantly less ($F_{1,11} = 16.007$, $p = 0.002$) average error than in the ROTATE condition (M = 6.6 cm, SD = 0.4).

The ANOVA did not find any significant interaction effects between the REFERENCE SYSTEM and BODY ROTATION variables.

To investigate closer we performed four post-hoc paired samples t-tests (and subsequently controlled for inflation of Type I error by using an adjusted α of 0.05÷4=0.0125). The first two post-hoc tests compared the HAND and NONE conditions separately in both BODY ROTATION conditions. When the participant did not rotate (i.e., in the STAY condition) there was no significant difference but when participants did rotate they were significantly more accurate when using their hand as a reference ($t_{11} = 4.621$, $p = 0.001$).

Similarly, the last two post-hoc tests compared the STAY and ROTATE conditions separately in each REFERENCE SYSTEM condition. When the hand was used as a reference system there was no significant difference in accuracy between the STAY and ROTATE conditions, however participants were significantly more accurate when not rotating in the NONE condition ($t_{11} = 3.978$, $p = 0.002$).

With respect to task completion time, the ANOVA also indicated that ROTATE trials took longer than STAY trials ($F_{1,11} = 42.135$, $p < 0.001$). This effect was completely expected as it took time for participants to perform the rotation.

In addition, the ANOVA also showed a significant overall main effect for the REFERENCE SYSTEM factor ($F_{1,11} = 5.657$, $p < 0.038$) on task completion time. Although the difference was small (8.3 sec. compared to 8.7 sec.) the hand condition was significantly slower than without the hand, presumably because of the extra time required to lift the non-dominant hand.

3.3.8 Discussion

Study 1 demonstrated that, even though users are able to create simple drawing, they struggled when multiple strokes are required and when their visuospatial memory was disrupted. In this study we confirmed that causing a disruption of visuospatial memory, in this case by forcing the participant to make a quarter turn, led to significantly lower accuracy. However, this effect was significantly lessened when participants used their non-dominant hand as a reference.

In other words, by allowing the user to associate their mid-air gestures with a movable reference point within their control (i.e., their non-dominant hand), the imaginary interface users provided the user with the necessary structure to reestablish the imaginary interaction plane after their visual field (and hence visuospatial memory) was disrupted.

However, we must still deal with situations when visuospatial memory has completely degraded or was never present in the first place. The following study addresses this.

3.4 POINTING BY COORDINATES

The previous two studies explored the use of Imaginary Interfaces in situations where users either create a simple sketch or recall a sketch from memory to annotate it. While we think of this sort of unstructured input as being a common application scenario, there are situations where users might want to take specific coordinate data into account when drawing or pointing. A user might construct a stock curve that passes through certain specific points or users might want to point to a coordinate pair received over an audio connection (the audio manual says "the green button is located at. . . ").

Here, we offer a very basic mechanism for Imaginary Interfaces to support this. As illustrated by Figure 3.12, when the user performs an 'L' shaped gesture with their non-dominant reference hand, a coordinate system is established. The user's thumb forms the unit vector along the x-axis and the index finger forms the unit vector along the y-axis. This coordinate system allows referring to locations on an imaginary interaction plane using coordinate pairs, such as (2,1) in the shown example. Negative coordinates can be used to refer to locations left of the index finger and below the thumb.

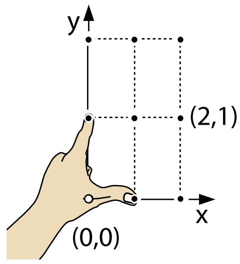


Figure 3.12: The user's thumb and index finger span a coordinate system that gives each point on the plane a unique address. Here the point "two thumbs right, one index finger up" is labeled (2,1).

To understand the capabilities and limitations of coordinate-based pointing we conducted a study that tested participants' ability to point to a location specified in Euclidian coordinates, more specifically vertical units of index fingers and horizontal units of thumbs. We found targeting error correlated with the Manhattan distance of the target from the user's fingertips, which serve as visual landmarks.

3.5 STUDY 3: POINTING BASED ON COORDINATES

In this study participants acquired targets given to them as coordinate pairs in (thumb, index) length units. We measured error as the Euclidean distance from the target. We hypothesized that error would grow with the distance from the tips of index finger and thumb, which participants use would as visual landmarks.

3.5.1 Task and apparatus

For each trial, participants started in a neutral position with their hands held loosely at their sides. Participants then received the target location as two digits via audio and displayed on a monitor. They pressed a footswitch, which started the timer and the trial.

Participants then raised their hands, performed an 'L' gesture with their left hand and acquired the indicated target by pinching with their right hand (Figure 3.13a). Participants committed the acquisition by pressing a foot switch again. This prompted the system to record the 3D location and rotation of the body and both hands, to play a confirmation sound and to complete the trial.

In preparation for the next trial, participants dropped their arms and rotated approximately 90° to simulate mobile use.

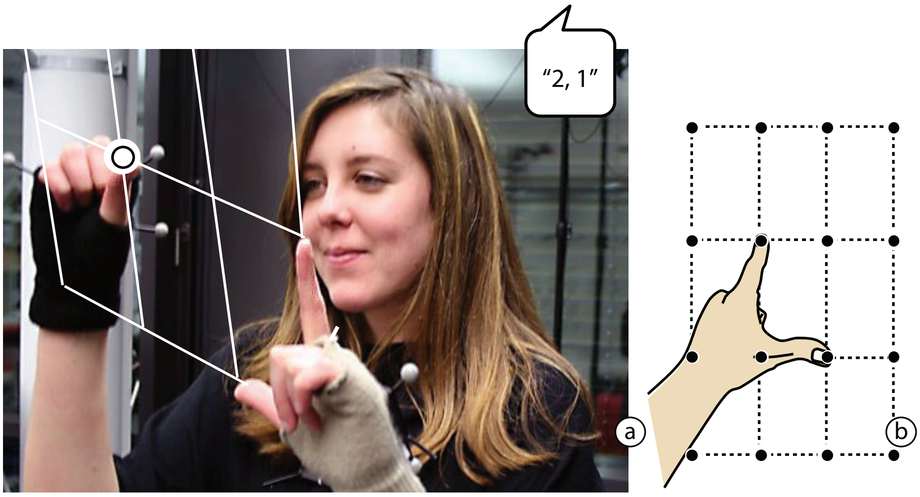


Figure 3.13: Study 3 task – (a) Participants selected a target at the coordinate announced by the system, (b) the coordinate was chosen from a set of 16 positions from (-1,-1) to (2,2).

We used the marker-based tracking system from Study 1 and Study 2 again as the apparatus for this study.

3.5.2 Experiment Design

Participants selected all targets on a 4×4 grid (Figure 3.13b) five times each in random order for a total of 80 trials. Together, the 12 participants completed 960 trials. All participants completed the trials in 30 minutes or less.

3.5.3 Participants

We recruited a new set of 12 participants (five female) from our institution and community. Two participants were left-handed but both used the mouse in their right hand and performed the experiment using their right hand. Participants were between 21 and 27 years old ($M = 23.0$, $SD = 2.0$). Each received a small gift for their time.

3.5.4 Hypotheses

Since the previous study showed that participants benefit from using their non-dominant hand to set the reference frame, we expected that error would increase with distance from this hand. In particular, we expected error to increase the further the selection point was from the two main landmarks, i.e., the fingertips of the 'L' hand:

HYPOTHESIS H1 Pointing accuracy will be highest at the fingertips.

HYPOTHESIS H2 Pointing accuracy will decrease as the distance from the nearest fingertip increases.

3.5.5 Data preparation

To allow us to compare selection positions between participants, we performed a more complex (compared to the previous two studies) set of data preparation operations that corrected for differences in participants' interaction planes and hand sizes.

First we corrected for participants interacting in planes of different tilt. For each participant, we determined the interaction plane using a linear planar regression based on all of their 3D selection positions. We then rotated this plane into the x-y plane and projected onto that plane, discarding the z coordinate. All further processing was done with the resulting 2D data.

Next, we corrected for differences in finger sizes. Since the tracking system tracked only entire hands, we reconstructed finger tips from the data for the (0,0), (0,1) and (1,0) targets. Following this, we corrected for the rotated and skewed interaction planes based on the three points gathered in the last step. The coordinate space was transformed such that the tip of the index finger was located at (0,1), the tip of the thumb at (1,0) and the origin at (0,0).

Note that the coordinate transformation and finger size calculation is a byproduct of our apparatus. It is not necessary for a camera-based prototype (such as we describe in Chapter 4) because such a prototype would actually "see" finger lengths and could establish the interaction plane directly from the fingers.

We removed 16 outlier trials (1.7%) from the data. We defined an outlier trial where either the x or y position was more than 1.5 times the interquartile range above the third quartile for each position after the above corrections. Many of the outliers appeared to be the participants mistakenly swapping the x and y coordinates.

3.5.6 Results

To summarize the selection accuracy across participants we calculated the minimum button size for each location that would capture 95% of the participants' selections. This metric, also known as error spread or *variable error* was calculated as four times the standard deviation (two standard deviations on each side of the mean) per user per location. As shown in Figure 3.14, we aggregated these data by the Manhattan distance from the nearest visual landmark (i.e., the index finger tip or thumb tip) for each coordinate the participants selected in the study. The figure clearly shows that the locations with a Manhattan distance of 0 (i.e., the finger tips) were indeed the most accurate and that minimum button size increased as the distance from the finger tips increased.

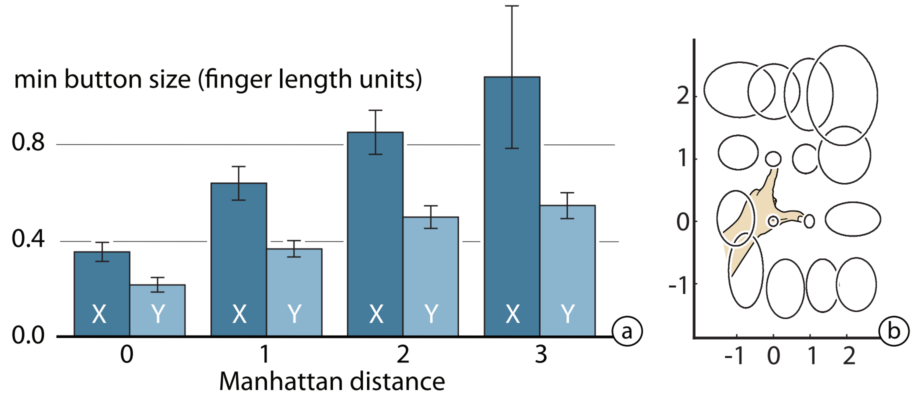


Figure 3.14: Study 3 aggregate results – (a) error spread by Manhattan distance from the nearest finger tip (± one standard error of the mean). (b) The same data overlaid as oval buttons on the imaginary interaction plane.

Breaking these results down per participant (shown in Figure 3.15), we see a fair amount of variation, especially at the extremities. Again, each oval encodes two standard deviations around the mean position (i.e., 95% of all targeting positions) for that location. You can see that some participants are consistent overall and others, while accurate close to the fingers are much less accurate in the space above the hand.

3.5.7 Discussion

In this study we quantified the positioning error for 16 locations surrounding the left hand and calculated the smallest button sizes that remain selectable at these locations.

As expected the fingertips were the most accurate locations, with button sizes of 0.35 thumbs wide and 0.21 index fingers high. As the Manhattan distance increased from the nearest finger tip error increased

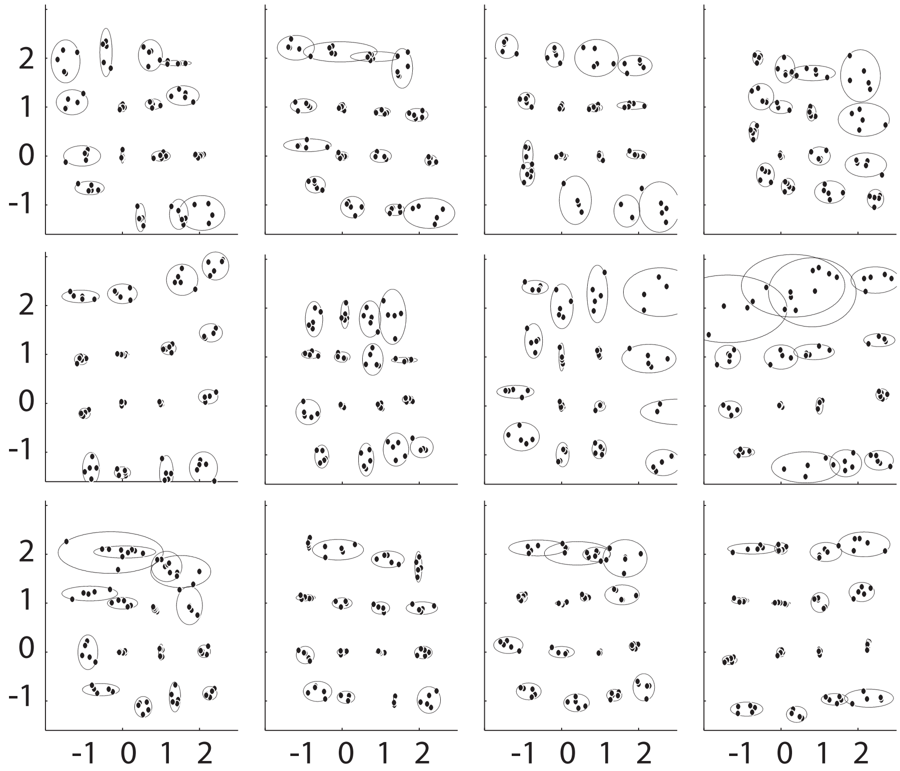


Figure 3.15: Study 3 raw results – error per target position per participant in finger length units. Each oval represents the minimum buttons sizes that would capture approximately 95% of selections for that target per participant.

linearly as shown in Figure 3.14. Overall, participants were able to select targets that were 0.722×0.401 finger lengths in size. Given that from the average index finger size is 10.66 cm (measured from thumb crotch to tip) and average thumb length is 6.66 cm (Greiner 1991) then the overall average minimum button size is 4.81 cm in the x direction and 4.27 cm in the y direction.

However, the size of the targeting areas differs greatly within the range tested and between participants. Upon inspection of Figure 3.15 you can see that some participants had a lot of trouble with the more distant targets while others did not. We expect that with better training all users could reasonably target distant imaginary targets.

Despite complaints from the participants, who stated they did not like the locations in the negative quadrants, these locations offered similar targeting ability as those in the positive quadrant.

3.6 SUMMARY OF THE STUDIES

Since Imaginary Interfaces rely, in part, on well-maintained visuospatial memory anything that restricts the creation or disrupts previous memory of the necessary spatial knowledge will affect how well people can operate these "feedbackless" interfaces.

The first study in the chapter showed that although small scale drawing could be created in a consistent manner, the multi-stroke drawing began to push the limits of what was possible with this simple version of Imaginary Interfaces. In the second study, we showed that moving within the environment also had a significant impact on the retention of the spatial memory necessary to retarget a position that had been previously drawn. However, performance could be significantly improved by using the non-dominant hand as a visual reference point. In the final study we showed how participants were able to select locations based on instructions encoded as a coordinate pair containing multiples of index finger and thumb length, and that selection accuracy diminished with distance from the finger tips.

With these studies we have gained first insights into the design space of Imaginary Interfaces, and can recommend a few basic guidelines to designers of imaginary interfaces:

1. Keep the required drawings simple. Users are readily able to produce basic shapes on an imaginary interaction plane but have problems as the complexity of the drawing increases.
2. To avoid overwhelming the user's visuospatial memory or allowing it to degrade, users should be encouraged to create annotations right away while memory is still fresh. By waiting too long, changes in scenery could make it more difficult to reestablish the imaginary interaction plane.
3. When mobile, allow the user to set the reference point with their non-dominant hand. This helps the user to reestablish their visuospatial memory by providing a strong visual reference points that demonstrably improves performance in reestablishing the imaginary interaction plane.
4. Exploit the features (such as finger length) of the reference hand to provide directions to imaginary interface elements. For example, "the OK button is located at position 1,2."
5. Since pointing accuracy decreases as the distance from the reference hand's fingers increases, allow users to draw and point close to their reference hand. To allow users to use the entire imaginary space effectively, complement small imaginary buttons close to the reference hand with large imaginary buttons at a distance.

CONCLUSIONS

This chapter provided an initial elaboration of Imaginary Interfaces, user interfaces that allow users to interact spatially without visual feedback. While traditional interfaces are either spatial (e.g., touchscreens) or non-visual (gestures), Imaginary Interfaces combine both aspects.

In this chapter, we showed that, despite offering no explicit visual or audio feedback, users of imaginary interfaces are able to produce simple drawing and annotate them, with the help of the non-dominant hand as a reference. As an initial investigation into the foundations of interactive Imaginary Interfaces, this chapter showed that users were able to select mid-air targets with reasonable accuracy, but with decreasing accuracy as the distance from the finger tips increases.

The studies in this chapter lay the foundation for Imaginary Interfaces, whose main purpose is to bring spatial interaction to screenless mobile devices. Such devices have the highest potential for miniaturization. The ability to create and share simple sketches on the fly opens up a range of new application scenarios, such as providing short gestural input or sharing sketches with driving directions as part of a phone call. Beyond this, as I will come to in later chapters, users could also interact with a conventional spatial interface that they previously acquired knowledge of, or with the liberal use of audio feedback, users are able to browse and operate completely unfamiliar imaginary interfaces.

The apparatus we built for these studies was not the final intended form factor of Imaginary Interfaces. We chose an optical tracking installation to be able to concentrate on the human performance that was possible with the interface and to avoid measuring only the quality of the prototype.

However, to validate the concept of Imaginary Interfaces I must also show a viable mobile form factor. In the following chapter, I present two hardware prototypes that support, respectively, the mid-air style interaction explored in this chapter and a new palm-based interaction. These prototypes, while not production ready, represent two possible sensing platforms that could be used to produce deployable imaginary interfaces.

4 FORM FACTORS AND PROTOTYPES

In this chapter I further develop the concept of Imaginary Interfaces by presenting the design of two different interaction styles, *mid-air* and *palm-based*, along with two hardware prototypes that implement these interaction styles. Both prototypes are based on a chest-mounted camera sensing platform that tracks the users hands while interacting with the interface.

As introduced in the previous chapter, the mid-air interaction style places the imaginary interaction plane in the empty space in front of the user (Figure 4.1a). By using their non-dominant hand to frame the interaction space with an 'L' gesture, users can select targets and draw using a pinch gesture with their dominant hand.

On the other hand, the palm-based interaction style, which I introduce in this chapter, places the imaginary interaction plane directly onto the palm of the user's non-dominant hand, as shown in Figure 4.1b. Users interact simply by tapping on the surface of their hand, just like a touchscreen. This readily available surface provides many tactile and visual landmarks that encourage learning of target locations.

After introducing the two interaction styles and prototypes, I conclude with a formal evaluation of selection accuracy using each interaction style. We found that the palm, presumably because of the abundance of visual and tactile features, allows for significantly higher selection accuracy (the palm-based interface was 36.6% more accurate than the mid-air interface).

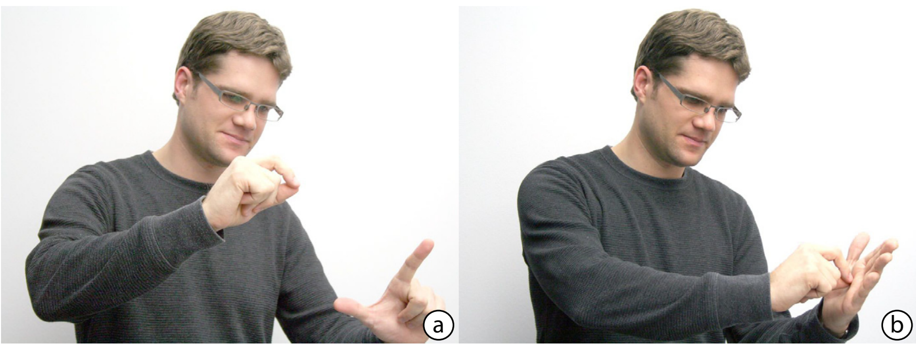


Figure 4.1: Interaction styles of Imaginary Interfaces – (a) *mid-air* interaction uses the empty space in front of the user as the interaction "surface", while (b) *palm-based* interaction exploits the user's non-dominant hand as the interaction surface.

Contributions

There are three main contributions in this chapter: 1) a further description of the mid-air interaction style and the implementation details of a representative prototype; 2) a description of the palm-based interaction style, also with a representative prototype; and 3) a formal comparison of the two interaction styles.

4.1 MID-AIR IMAGINARY INTERFACES

The mid-air interaction style, as introduced in the previous chapter, uses the empty space in front of the user as the interaction surface. The user establishes the lower-right corner of the imaginary interaction plane by forming an 'L' gesture at that location, then pinches in space to provide input to the imaginary interface.

Since there is no visual feedback from an imaginary interface, we were free to choose the most user-accessible space for the imaginary interaction plane: the empty space directly in front of them. It is a large interaction space, that can be expanded at will, as long as its bounds remain within the reach of the user.

The 'L' gesture sets both the origin and the unit size of a 2D coordinate system. The origin is located at the joint between the index finger and thumb, and the unit lengths for the x and y axes are set by the length of the thumb and index finger, respectively. This visual reference allows users to acquire targets using coordinates of the style "two thumbs up and three index fingers to the right". In the previous chapter, I showed how this allowed for reliable acquisition of targets measuring on average 4.8×4.3 cm.

To make selections and to draw on the imaginary interaction plane, users employ a pinch gesture. The pinch gesture is relatively simple to sense with computer vision techniques (as we described in the following section) and also has a distinct benefit for interaction, as described by Wilson (2006): there is perfect correspondence between the user feeling the tactile feedback of there thumb and index finger touching and the computer system sensing that a pinch has occurred. Compare this to a pointing gesture based on piercing an imaginary interaction plane: the user receives no immediate feedback from the environment when they 'touch' the interaction surface because there is nothing physical there to touch.

The 'L' gesture has another benefit. It can provide a unique 'on' gesture for an interactive system based on the imaginary interface. If the system does not see a hand forming the 'L' gesture, all input could be ignored. Once it observes the gesture, the interactive features of the imaginary interface could be instantly invoked. Since the gesture is an unusual posture, it is unlikely that users would accidentally invoke the system and perform unwanted input.

4.1.1 Mid-air prototype

For the benefits of Imaginary Interfaces to be realized, the sensing mechanism must be small and mobile. In the user studies in the previous chapter we used an fixed optical tracker installation to prototype interaction with an imaginary interface. While practical for a research installation, that setup requires an expensive and fixed installation in the target environment and is therefore not representative of how an imaginary interface might be implemented in the real world.

A wide variety of sensing mechanisms could be used to track a user's hands and could thus be used to implement a mid-air imaginary interfaces (see Section 2.3 for a survey of sensing technologies). Inspired by the Gesture Pendant (Starner et al. 2000), we chose a chest-mounted camera with IR illumination to sense the user's hands as they interact in front of the user. Our prototype, shown in Figure 4.2, tracks the hands of the user when held in front of their chest using computer vision techniques.

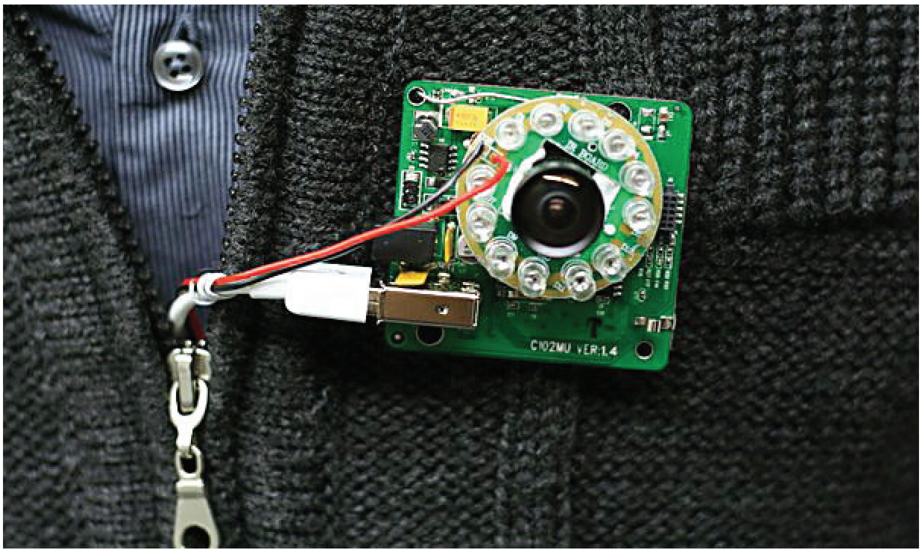


Figure 4.2: Closeup of the mid-air prototype mounted to a user's chest.

The device consists of a Fire-i black-and-white camera with a 107° wide angle lens equipped with an IR-pass filter. The camera returns a 640×480 pixel grayscale image at 30 Hz representing the scene in the IR spectrum. Surrounding the lens is a ring of IR LEDs that illuminate the user's hands. The camera is connected to a computational backend (on a laptop computer) via Firewire 400. Overall, the sensing components measure 5.25×5.25 cm.

The wide field of view of the camera used in this prototype allows the user substantial freedom regarding where to place their hands while interacting. However, there is a trade off. The wider the field of view, the less pixels will be used to represent the hands, resulting in lower sensing accuracy. We found the lens field of view in this prototype was appropriate for our purpose of showcasing the technology, but a deployed solution might consider a higher resolution camera or perhaps an actuated camera that follows the hands, as recommended by Mayol-Cuevas et al. (2009).

Our prototype obtains good separation of the user's hands from the background by illuminating the space with infrared light (Figure 4.3a). Since the hands are close to the illuminant, they are much brighter than the background scene. Note, however, that this approach requires controlled lighting and, for example, does not work in sunlight, which would overwhelm the relatively weak illuminant.

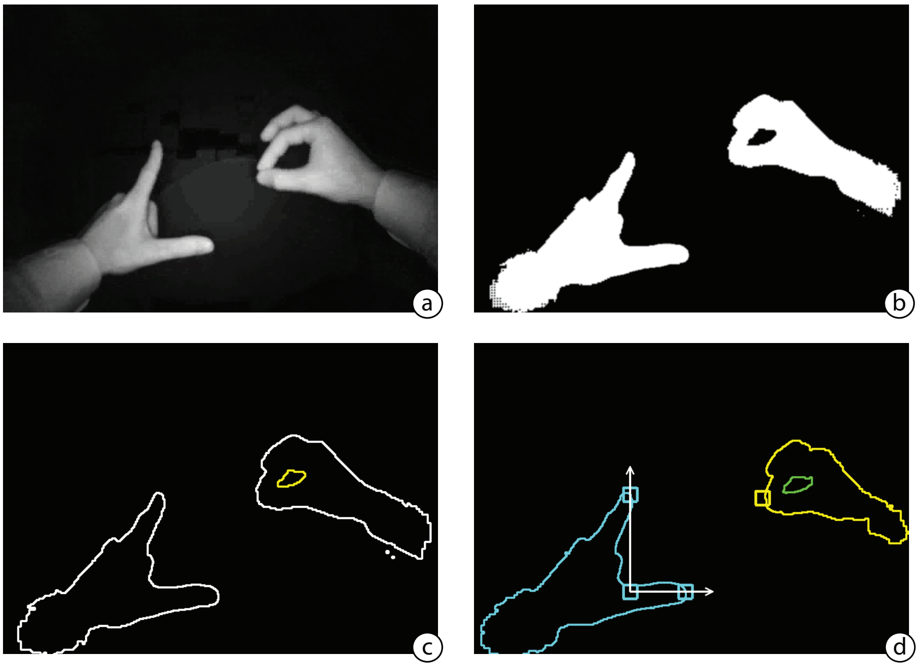


Figure 4.3: Computer vision pipeline for the mid-air prototype: (a) Retrieve raw image from the camera; (b) apply threshold; (c) find contours; (d) determine three points of the 'L' gesture from left hand and the pinch point from the right hand.

The computer vision pipeline can be broken down into these steps: First, the raw image is thresholded (Figure 4.3b) separating the hands from the background. From this binarized image the system finds all contours (Figure 4.3c) and discards all but the two largest groups (designated left and right hands). Then, as shown in Figure 4.3d, the system determines the tip of the thumb and index finger by finding the top-most and right-most points and uses these points to define the interaction plane.

In Figure 4.3d the pinch location is also shown. Employing the approach of Wilson (2006), the system determines that a pinch has occurred by testing for the presence of a interior contour in the pinching hand. An unpinched hand will not have this interior contour, providing a robust mechanism of determine if the user has pinched. The system determines the pinch location to be the left-most point on the pinching hand. Finally, this pinch location is transformed to fit the coordinate system defined by the left hand and passed as a 2D coordinate point to the currently running application.

4.2 PALM-BASED IMAGINARY INTERFACES

A mid-air imaginary interface lets users point to an interaction surface located in empty space in the space in front of them. By framing the interaction area with an 'L' gesture (Figure 4.4a) users were able to redefine the imaginary interaction plane where ever and whenever they would like. Alternatively, there is another convenient location for an imaginary interaction plane that is just as convenient to use: the palm of the non-dominant hand (Figure 4.4b).

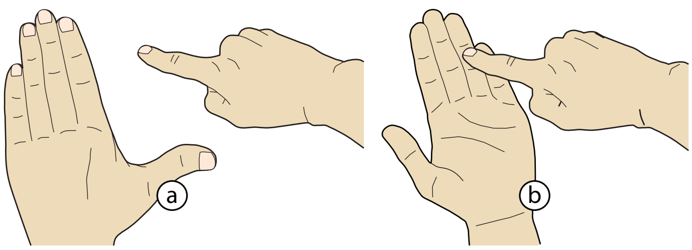


Figure 4.4: (a) With a mid-air imaginary interface users interact in empty space framed by the 'L' gesture, (b) with a palm-based imaginary interface, the interaction surface coincides with the surface of the non-dominant hand.

While a mid-air imaginary interface offers a much larger interaction area, the palm-based version offers a benefit that could be very beneficial when learning an imaginary interface: *memorable landmarks*. As I showed in Section 3.5, proximity to landmarks—in that case the tip of index finger and thumb—helped participants to acquire targets. However, the empty space of the a mid-air imaginary interaction plane is all but devoid of landmarks. The palm, in contrast, is full of landmarks, many of which even have commonly known names (e.g., the index finger tip), allowing users to create symbolic associations.

As a side effect, on the palm, a tap is established by the physical contact, very much like on any touch screen. This results in four additional benefits:

1. *Stabilize the finger:* physical contact between hands stabilizes the finger during pointing, perhaps resulting in less fatigue (Park et al. 2012).
2. *Eliminate pinching:* most users are more experienced and thus skilled with tapping than with the pinching gesture required by a mid-air imaginary interface.
3. *Spatial haptic feedback:* during tapping, the sensation on the non-dominant hand reflects the acquired location, providing an additional cue for target location.
4. *Eliminate parallax:* when targeting in empty space, the finger is free to move in 3D and is thus often outside the 2D interaction plane of the imaginary interface. Mapping the finger position to the desired 2D point on the plane, however, is subject to ambiguity and pointing error because we cannot know how the user conceptualizes this projection: orthogonal projection, line-of-sight, etc. Pointing on the palm avoids this problem.

While it is hard to compare the overall number of addressable locations in empty space to the palm, the combination of the four factors listed above increases the pointing resolution on the palm. Also, the small size of the palm-based interaction surface could be less socially awkward to use in public than the large size of the mid-air interface.

4.2.1 Palm-based prototype

Sensing hardware

Interaction on the palm opens up other sensing options, such as gloves (Kuester et al. 2005) for example. However to retain the natural feel of an empty hand, we chose to use a vision-based approach like the mid-air prototype. Unfortunately, to sense touch on an uninstrumented hand, the IR camera based approach of the previous prototype will not work. When the two hands overlap they will appears as one large 'blob' to the computer vision pipeline. There is not enough information to be able to separate the hands.

Therefore, we chose to use a time-of-flight depth camera (see Kolb et al. (2010) for detailed discussion of their abilities) because, unlike other approaches, the extra depth information allows us to separate the hands when they are overlapping. Figure 4.5 shows the prototype in action.

Figure 4.5: The prototype tracks input using a time-of-flight depth camera (PMD[vision] CamCube).

The time-of-flight camera also enables our prototype to work in all lighting conditions, including outside in direct sunlight (as shown in Figure 4.6), unlike standard infrared cameras or other types of depth sensing cameras (for instance Microsoft's Kinect camera). Our depth camera is a PMD[vision] CamCube that provides frames at 40 Hz with 200×200 pixel resolution.

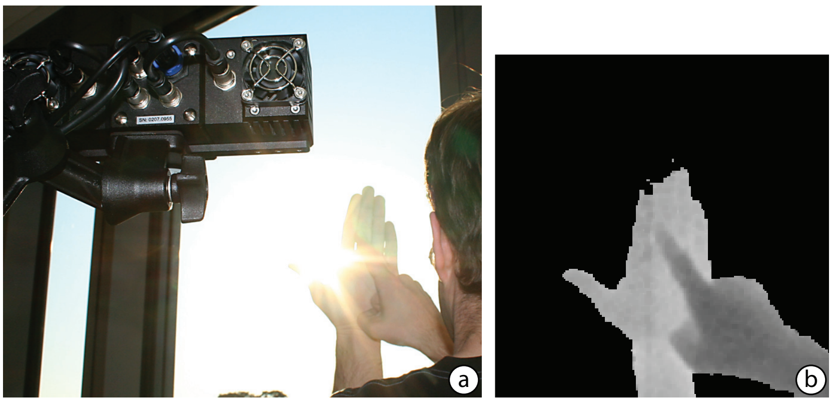


Figure 4.6: (a) The time-of-flight depth camera even works in direct sunlight. (b) The unaffected output of the camera.

Although we mounted our camera on a tripod looking over the user's shoulder (our camera was large and heavy, see Figure 4.5), depth cameras have evolved to be small enough to be mounted on the chest just like the infrared camera from our first prototype. This is an active area of development and recent offerings, such as the CamBoard mini (Figure 4.7b) and nano (Figure 4.7c), are becoming much smaller. This rapid miniaturization suggests that it might be possible in the near future to have robust depth sensing from very small wearable components. If this happens, our vision of a simple chest-mounted sensing mechanism would be possible.

Figure 4.7: The line of PMD[vision] time-of-flight depth cameras: (a) CamCube; (b) CamBoard mini; (c) CamBoard nano.

Algorithm

In order to extract the two hands from the input image, we process the raw depth image as shown in Figure 4.8. From the raw depth map, we find the closest pixels, remove all pixels with relative depth values of more than 30 cm from the closest and smooth all remaining values. To determine the number of visible hands, we create a histogram of depth values in the masked image and calculate the number of strong peaks (indicated by green squares in Figure 4.8c). We classify the two hands based on the two distributions in the histogram but splitting the histogram at the lowest point between the two peaks. From this classification we obtain the masks for the pointing hand and the reference hand's palm.

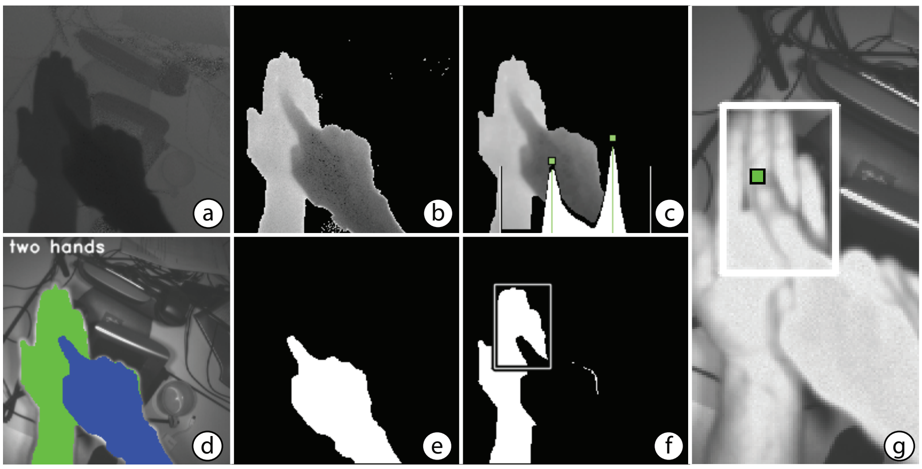


Figure 4.8: In processing the (a) raw depth image, our system (b) thresholds and (c) calculates a depth histogram to (d) segment the image into two masks: (e) pointing hand and (f) reference hand. From that we calculate (g) the final touch position and reference frame.

To determine if and where the user is touching the palm, we pick a location inside the pointing hand mask and fill using a small tolerance value, eventually walking "down" the finger towards the reference hand. If the fill does not reach a depth value that belongs to the reference hand while staying within the tolerance value, we infer that the user is not touching. If it does, we infer that the user's finger is touching the palm.

Due to the limited resolution of the depth camera, we cannot find the precise end of the touching finger. Instead, we determine the touch location from the end of the point mask offset by a small vector in the direction of the finger (green square in Figure 4.8g).

The width of the reference frame for touch events is set to the width of the fingers excluding the thumb. First, we calculate the width of the palm 3 cm from the top of the hand to exclude the thumb, and we set the height of this reference frame to match an aspect ratio of 1.5. The final frame is shown in Figure 4.8f and g. As this reference frame is subject to noise if the pointing hand is present, we update the reference frame size only if one hand is visible and, upon sensing both hands, only translate the reference frame's location.

As the computed raw locations are subject to strong noise, we use hysteresis to maintain touch states (touch/no touch) and to smooth input coordinates, which enables smooth dragging or even free-form drawing. This also prevents processing inadvertent input, such as a hand waving in front of the camera. Our system supports all of the same single-touch interactions that are possible on the phone: swiping, scrolling, tapping, dragging, drawing, etc.

4.3 STUDY: TARGETING IN MID-AIR VS. THE PALM

The two form factors for Imaginary Interfaces presented in this chapter have different characteristics. The mid-air interface has a large interaction surface with a builtin coordinate system based on finger and thumb lengths, whereas the palm-based imaginary interface has a fixed, relatively small size but with an abundance of visual and tactile features that we expect would improve performance.

In this section, I present a formal lab study that compared the performance of these two interaction modes.

In order to evaluate the interaction style rather than the current condition of our prototype (whose touch resolution is limited by the depth camera's 200×200 pixel resolution) we conducted this study using "perfect" tracking, i.e., post-hoc analysis of high resolution photos.

In this study we chose randomly placed targets on the participants' hand, that is, the targets were not directly aligned with finger segments or other features on the hand. Therefore these results show performance worse than with a layout specifically designed to align widgets with landmarks on the hand.

4.3.1 Interfaces compared in this study

In the MID-AIR condition, participants targeted in the space framed by their thumb and index finger (Figure 4.9a) and in the PALM condition participants targeted on palm of their hand (Figure 4.9b). The size of the tracked area was kept constant for both conditions.

To shorten training, we used the *wallpaper* approach to be described in the following chapter (see Figure 5.9) where participants' own hand was displayed behind the targets. With this approach, they were able to readily associate the arbitrary target locations with landmarks on their own hand.

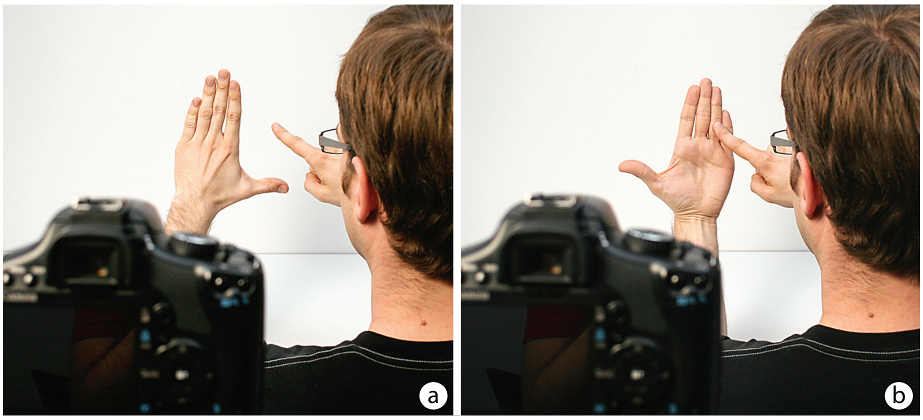


Figure 4.9: Study apparatus – (a) MID-AIR and (b) PALM interface conditions showing the camera used to record the interaction position.

4.3.2 Task and procedure

During each trial, participants selected three locations at a time (in pilots we determined participants were able to remember three locations easily). Participants learned the three target locations by repeatedly targeting them on a screen device (here an iPod Touch) until they were able to reliably target with at least 5 mm accuracy on the touch screen, as shown in Figure 4.10a and b. Participants were then prompted repeatedly with a target number and responded by selecting the respective position (Figure 4.10c) in MID-AIR or on their PALM, depending on the condition.

There were two independent variables: Target Location (4 groups of 3 locations) and Interface (MID-AIR vs. PALM). As a within-subjects de-

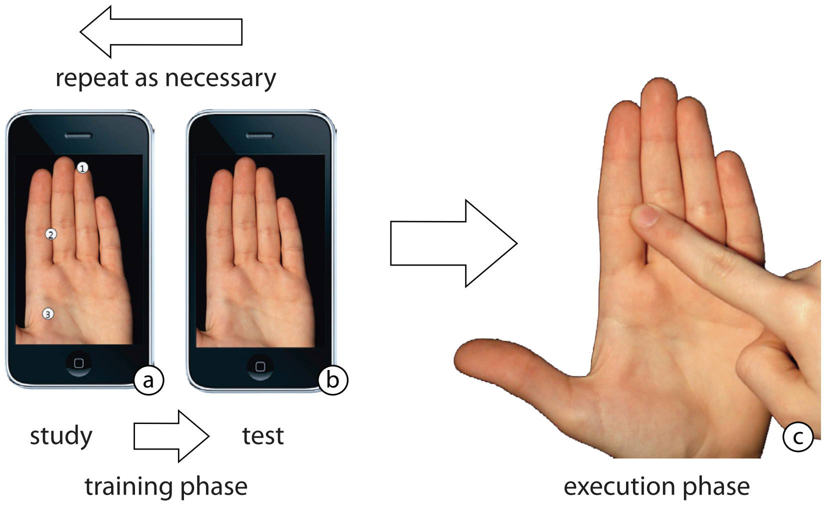


Figure 4.10: Study task – (a) During the training phase, participants learned target locations by selecting them on an image of a hand. (b) The participant was then required to reproduce those locations without feedback to ensure they successfully learned the locations. (c) During the execution phase, participants selected those locations using the imaginary interface for the condition.

sign, half of the blocks used the MID-AIR interface and the other half the PALM interface. The order of interfaces was counterbalanced across participants. Participants recalled and touched each target five times in random order; each participant completed four such blocks (two for each interface) with each block featuring a different set of 3 targets. Together, the experiment consisted of 12 participants $\times$ 2 interfaces $\times$ 2 blocks $\times$ 3 targets $\times$ 5 reps = 720 trials. Finally, participants filled out a short questionnaire indicating their preferred interaction surface (reproduced in Appendix B.1). Each participant completed the experiment session within 30 minutes.

4.3.3 Apparatus

During the training phase, participants used an iPod Touch running custom software that provided the training stimuli.

During the execution phase, we used a DSLR camera to record participants' touch interactions (Figure 4.9). When the participant pressed a foot switch, the experiment application triggered the camera to take a photo. After the experiment was complete, we manually extracted touch locations from the high-resolution photos on a millimeter level, which kept tracking errors to a minimum.

4.3.4 Participants

We recruited 12 participants (one female) from our institution. They were between the ages of 19 and 28 ($M = 21.8$, $SD = 2.56$). All participants were right-handed. They were given a small gift for their time.

4.3.5 Hypothesis

Reflecting our earlier discussion on the properties of the palm interface, we hypothesized that the PALM interface would allow participants to target with higher accuracy than in the MID-AIR condition. Furthermore, we were interested in quantifying the accuracy for these interfaces. Such information would be useful for system designers when choosing how large and close they would make the UI elements in the interface.

4.3.6 Results

For each trial, the study apparatus took a high resolution photograph of the participant's hands selecting the target. These photos (a selection is shown in Figure 4.11 and Figure 4.12) were manually analyzed after the experiment was complete to determine the touch position. We discarded 7 bad trials where no data was recorded and 15 outlier trials where the touch location was greater than three standard deviations away from the centroid of selections for that location. This left 698 trials for this analysis.

To allow for comparison between participants, we normalized all hand sizes so that the index finger was 7.25 cm long (the average of the population's male and female average index finger lengths (Greiner 1991)).

For this analysis we decomposed selection error into *offset* and *minimum button size*. These metrics are also known classically as *constant error* and *variable error*, respectively (Guth 1990). Offset is the distance from the target to the center of the selections made by participants for that target. It indicates a mismatch between where participants tend to select the target and its actual location. Minimum button size is a measure of the variability in their responses (i.e., noise), and represents the diameter of a button that would have captured 95% of all selections by this participant for this target. The minimum button size was calculated assuming one overall offset per target, instead of per-participant offsets. This is a liberal estimate of touch accuracy, because it is not calibrated per participant.

Figure 4.11: Study raw photo composites for *mid-air* condition from four participants (columns) for the first three targets (rows). The repeated trials from each participant are shown overlaid together in each subfigure.

Figure 4.12: Study raw photo composites for *palm* condition for the same four participants.

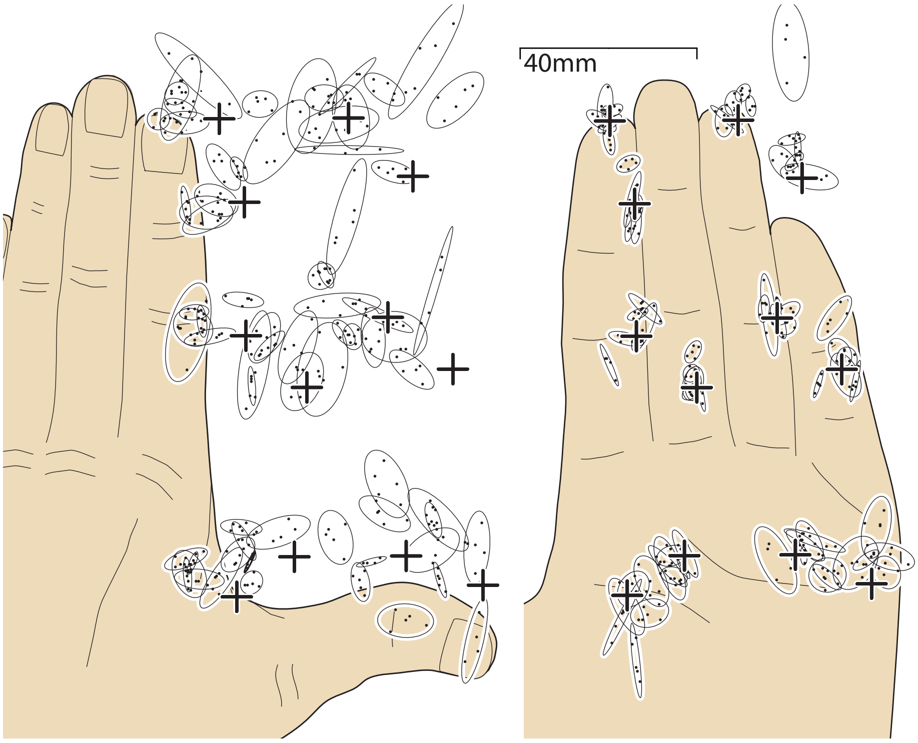


Figure 4.13: Study results – all touches from all participants for (left) the MID-AIR condition and (right) the PALM condition. Plus signs indicate actual target positions. Ovals represent the bivariate normal distribution of selections per participant per target.

Figure 4.13 shows all of the selection points grouped by target and by participant. Each ellipse shows plus/minus two standard deviations around the mean calculated from the bivariate normal distribution from the five trials from one participant for one target—it represents the per user minimum button size.

Overall the average offset for selections, shown in Figure 4.14a, was 8.556 mm (SD = 0.93) in MID-AIR and 4.65 mm (SD = 0.56) on the PALM.

The average diameter of a circular minimum button (calculated by taking the largest of the x and y diameter from the ovals), shown in Figure 4.14b, for the MID-AIR interface is 27.9 mm (SD = 3.16) and 17.7 mm (SD = 2.23) for the PALM interface, which is 36.6% more accurate. This difference was statistically significant ($t_{11} = 2.912$, $p = 0.014$, Cohen's $d = 0.84$).

All participants but one stated that they preferred the PALM to MID-AIR (one was undecided, stating that the mid-air interface would be best suited for drawing interactions, but the palm would be more appropriate for most other applications). Participants consistently remarked that they preferred the PALM because it offered "marks", "shapes", "spots", etc. that allowed them to orient better. For instance, one participant stated that "The palm is better because there are shapes that I

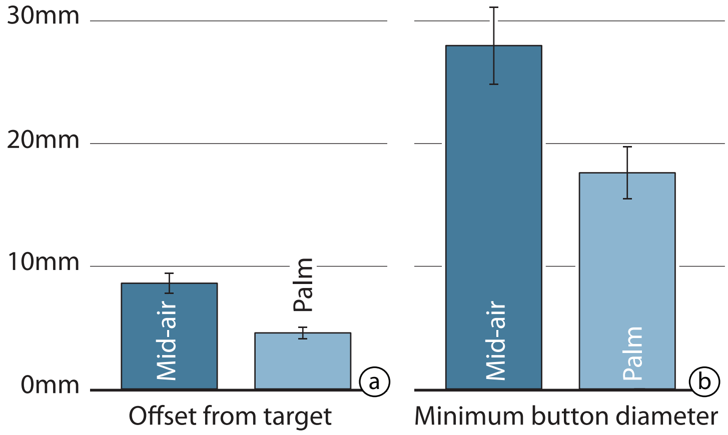


Figure 4.14: Study aggregate results – (a) average offset from target and (b) average minimum button diameter for the MID-AIR and PALM conditions. Error bars are ± one standard error of the mean.

can remember, [which is] easier to remember than distance in the air." Familiarity with the palm also seems to play a role. Another participant remarked "I know the palm of my hand because I have seen it a million times."

4.3.7 Discussion

As hypothesized, the PALM interface was significantly more accurate than the MID-AIR interface. Furthermore, the participants overwhelmingly preferred the palm-based imaginary interface. It seems that the extra features available on the PALM encouraged higher accuracy and that the empty space of the MID-AIR interface did not provide enough reference to consistently select the targets.

Therefore, we believe that the palm is an overall more appropriate interaction style for an imaginary interface and we recommend that system designers take this into account when building systems similar to Imaginary Interfaces.

That said, the specific quantitative results presented in this section were obtained with a tracking mechanism (i.e., post hoc analysis of high resolution photographs) more accurate than what our current prototype, or any known prototype, can deliver. Consequently, these values presented here should be considered a theoretical minimum for feedbackless targeting in mid-air and on the palm.

4.4 FUTURE FORM FACTORS

As cameras, processors, and wireless communication components continue to shrink and vision-based sensing continues to get more powerful, we envision future versions of sensing and processing components of Imaginary Interfaces to shrink to the size of a button (similar to Hännikäinen et al.'s (2005) button) or a necklace pendant (Starner et al. 2000). If so, they could be worn almost invisibly as part of the user's clothing as shown here in Figure 4.15. With the PMD CamBoard nano (shown in Figure 4.7c) the sensing portion is almost there today, however it requires a standard computer to process the images.

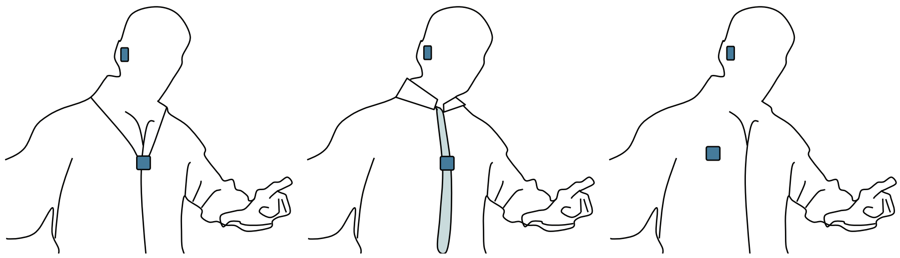


Figure 4.15: We envision future versions to be a small self-contained clip with cellular radio that can be worn as a brooch, pendant or clipped onto clothing.

As the sensing technology improves and electrical components continue to shrink, new types of interaction could emerge. We think of the two interaction styles presented in this chapter as only an initial investigation. We imagine that there are situations where other parts of the body might be useful as an interaction surface. For instance, while sitting on a train or a bus, the surface of the user's lap would be a convenient surface to perform input. These ideas, which have been well explored in the wearable computing literature (for instance, Rekimoto's (2001) GesturePad effectively provides a touchpad embedded in the user's clothing), could be enhanced by mapping Imaginary Interfaces style interaction on top of them.

We also believe that Imaginary Interfaces could be useful in non-mobile environments. To begin to investigate this we created an initial prototype based on a living room scenario. In this prototype, we mounted a Microsoft Kinect depth camera to the ceiling above a living room couch and determined touch location on the uneven touch surface with the algorithm introduced by Wilson (2010).

The resulting system, shown in Figure 4.16, which we called *draw-your-own* imaginary interface allowed the user to draw buttons, sliders and other elements with invisible ink on arbitrary surfaces and then interact with them as if they were real.

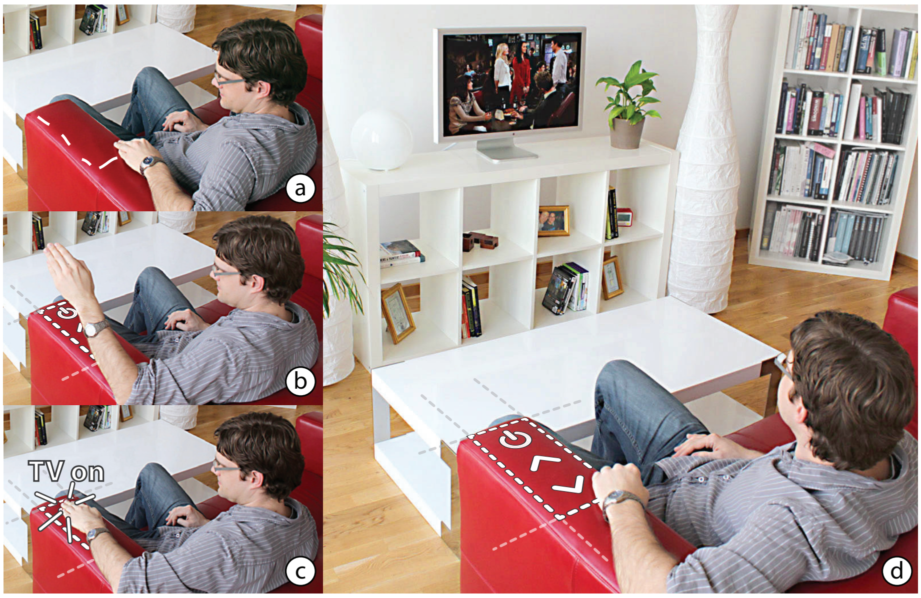


Figure 4.16: The principles of Imaginary Interfaces could also be applied to situations other than mobile interaction. Here a user is interacting with a *draw-your-own* imaginary interface on the couch in his living room. (a) He first draws a rectangle on the arm rest to tell the system where he wants the (b) remote control to be placed. (c) Then he can turn on the TV by touching the location associated with the ON button and (d) watch TV without looking for the remote control.

Although there is more research needed to confirm this concept, we believe that Imaginary Interfaces could transfer to other domains (such as the living room scenario presented here). However, the main challenge in all scenarios of use, including mobile, is to provide a mechanism to help the user learn the interface. In the following chapter, I present one technique to do this based on transferring the knowledge a user has obtained with a visual interface to operate a new imaginary interface.

5 LEARNING BY TRANSFER

In the previous chapters I showed how Imaginary Interfaces enable spatial input on mobile devices without requiring a screen. With their hands tracked by a chest-worn camera, users of imaginary interfaces point and draw in the empty space in front of them or on the palm of their hands. In Chapter 4, I showed that users are able to successfully select targets with both of these form factors but only when the user already knows where the targets are located. The question remains how a user would learn the precise location of each target on the imaginary interaction plane in the first place.

In the initial investigation into Imaginary Interfaces (in Section 3.5), I presented a naïve method of identifying target locations that uses spoken or written instructions to indicate the coordinates of UI elements. For a mid-air imaginary interface this meant describing locations using multiples of the thumb and index finger lengths, as in "for Mail, select one thumb right and two index fingers up." Unfortunately, this approach quickly becomes unwieldy with more than a few potential targets; the instructions quickly become reminiscent of a voice menu with a long list of potential options, a slow and frustrating experience (Yin and Zhai 2006). With practice, users would eventually be able to select targets without listening to the choices (i.e., to *dial ahead* (Perugini et al. 2007)), but since real-world interfaces can hold dozens of widgets, learning all the widget locations can take a long time, leaving users stuck with the inefficient voice menu style of interaction.

In this chapter, I present an alternative approach called *transfer learning* that addresses this difficulty. By designing imaginary interfaces that mimic the layout of mobile devices that users are already familiar with, we allow users to operate an imaginary interface by *mimicking* their use of the corresponding real-world screen interface (as shown in Figure 5.1). As I demonstrate in this chapter, users are able to take the spatial knowledge acquired when using a physical device and apply it to a corresponding imaginary interface. I will refer to this by saying that the spatial knowledge was *transferred* from the physical device to the imaginary interface.

Contribution

This chapter has one main contribution: the concept of learning imaginary interfaces by *transfer*. To showcase this concept I present the *Imag-*

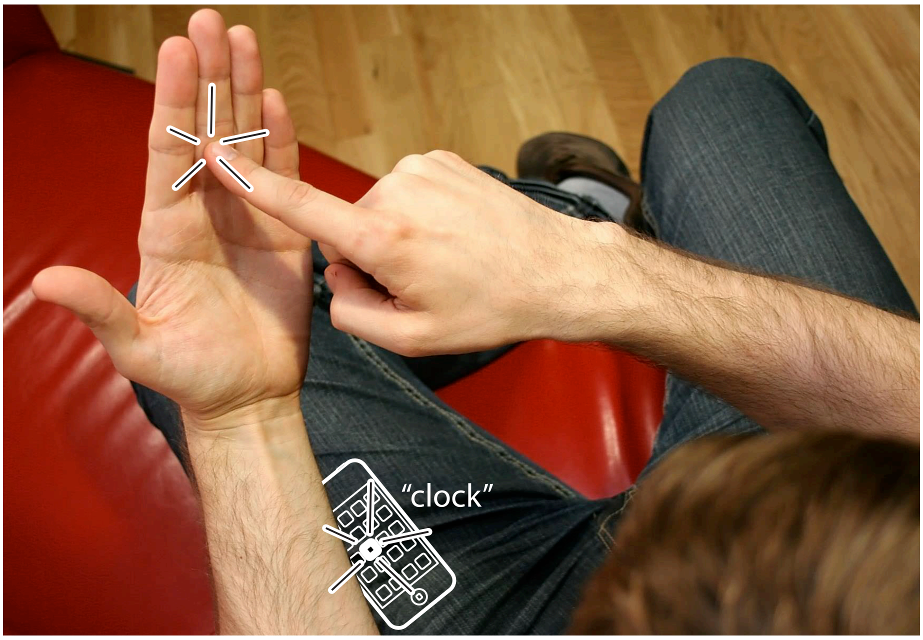


Figure 5.1: With the Imaginary Phone, this user controls the mobile phone in his pocket by mimicking the interaction on the palm of his non-dominant hand. The palm becomes a surrogate that can be used in place of the actual phone. Our prototype tracks the interaction and send touch events to the actual physical device to trigger the corresponding function. The user thus leverages spatial memory built up while using the screen device.

inary Phone prototype. Derived from the palm-based imaginary interface prototype described in Section 4.2, it allows users to use their palm as a shortcut interface for their phone. I present some design choices for the Imaginary Phone and validate the concept of transfer learning with two studies that show how participants are able to 1) readily learn the location of UI elements through everyday use; 2) transfer that knowledge to their palm; and 3) target on their palm with sufficient accuracy to select standard iPhone widgets.

5.1 THE IMAGINARY PHONE

I start by illustrating the concept of transfer learning with a system that we call the *Imaginary Phone*: an imaginary interface that offers a shortcut interface for an iPhone (Figure 5.1). Instead of retrieving and operating the physical phone, users *mimic* the interaction by pointing and dragging on their empty palm.

5.1.1 Prototype

Our prototype, derived from the palm-based imaginary interface from Section 4.2, tracks the pointing interaction between the user's two hands and sends the touch position to a physical mobile device, here an iPhone located in the user's pocket. The physical device supplies feedback to operations via the built-in speaker or a wireless earpiece worn by the user.

As described in Section 4.2, the system determines that the position of touch within the palm's interaction surface using a depth camera. The Imaginary Phone prototype then smoothes the touch position and relays it to the user's iPhone located in their pocket, as shown in Figure 5.2. A custom-written input daemon on the iPhone receives the touch events as TUIO packets sent over WLAN and creates touch events on the phone. Since the touch events are injected directly into the main operating system level input stream, all iPhone applications can be operated with the Imaginary Phone prototype. There is no need to write custom applications to handle the touch events originating from the palm.

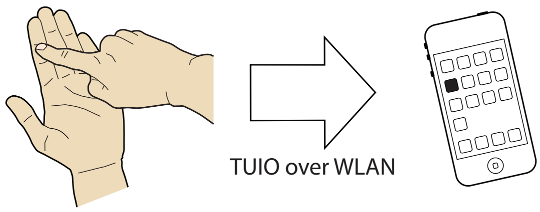


Figure 5.2: Once the system determines the touch position, it is forwarded in a TUIO packet over WLAN to the user's phone.

To supply feedback to the user, the accessibility mode of the phone can be used to provide auditory confirmation of actions.

Just like on the iPhone, the unlock gesture is required to unlock the Imaginary Phone. This allows the system to disregard spurious input that happens naturally when not using the system.

5.1.2 Walkthrough

Users can choose, either because it is necessary or just convenient, to use their Imaginary Phone for various quick tasks instead of retrieving the physical device from their pocket. Here is an example scenario:

A representative user is cleaning up the dishes and receives a phone call. Since his hands are wet, he cannot take the call on his physical phone and uses the Imaginary Phone with his wireless earpiece instead. He answers the call by swiping on his hand, which is the same interaction he would have performed on the physical phone. The call is from a friend that wants to go jogging tomorrow morning. The user agrees and ends the call by touching the location of the END button on this wet palm. He then launches the CLOCK application, selects the location for ALARMS and taps the toggle button to enable his early morning alarm to ensure he gets up on time.

Later, while watching TV, the user wants to order food but cannot find his phone in his pockets. Not wanting to get up from the couch and search, he chooses the Imaginary Phone to place a call as shown in Figure 5.3.

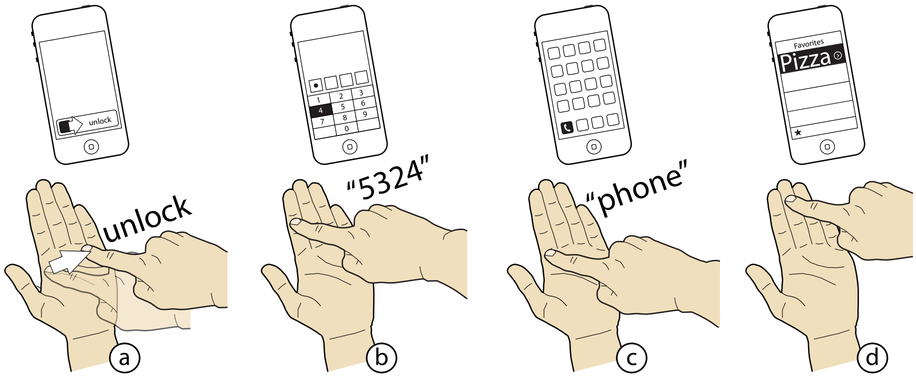


Figure 5.3: Walkthrough of making a call with the Imaginary Phone: (a) unlock with a swipe, (b) enter your pin, (c) select the PHONE function and (d) select the first entry from the speed dial list.

Throughout this walkthrough, the user operated the Imaginary Phone instead of retrieving his phone, either because it was impractical (dirty hands) or inconvenient (lazy on the couch). Since the Imaginary Phone is always available and can be operated without holding a physical device, it allowed the user to quickly perform the interaction and continue with his current activity.

5.1.3 Resulting interaction model and benefits

Interaction with an imaginary interface that is learned by transfer, such as by the user in the previous scenario, is possible only because the user has been using the physical screen device over a period of time and consequently learned the spatial locations of the necessary UI elements. This happens inadvertently—without extra effort users become increasingly familiar with the locations of such widgets over time. The spatial knowledge they gained from using the physical device can then be *transferred* to an imaginary interface.

The acquisition of this knowledge is imperfect and gradual but at some point users have performed an operation often enough to confidently know the locations and sequence of touches needed to execute it and they can begin to perform that operation on the imaginary version of the interface. This would occur one operation at a time, with the simpler and more common operations being transferred earlier. Therefore, microinteractions (Ashbrook 2010) will generally be the first to transition to an imaginary interface.

As a result of this gradual learning, transfer learning essentially turns the screen device into a *training mode* for the imaginary interface—or, depending on your perspective, the imaginary interface into an *expert mode* for the screen device. Accordingly, the benefit of the transfer model depends on the use case:

For users of physical devices, the main benefit of transfer learning is that it allows mobile phone users to perform some of their interactions on an imaginary interface instead. This saves these users the effort of retrieving the physical device, which, for short interactions, can make up a large proportion of the interaction time (Ashbrook, Clawson, et al. 2008). This speedup makes the Imaginary Phone particularly valuable for microinteractions, such as dismissing an alarm dialog, as they are performed regularly but do not last more than a couple seconds. Since transfer learning allows users to leverage their experience with the physical device, users can redeem these benefits right away, without the need for a separate training period.

For users of imaginary interfaces (i.e., users that only have access to the imaginary interface), transfer learning replaces the voice menu-style training period. Offloading the learning phase to a screen device (1) allows learning to take place in a visual and inherently parallel way and, (2) unlike when using a voice menu-style interface, interaction is fast during training, lowering the entrance barrier to learning imaginary interfaces.

5.1.4 Making it work: the three requirements

Transfer learning with the Imaginary Phone can be broken down into a chain of three logical steps (Figure 5.4), each of which depends on one assumption:

1. *Spatial memory:* while using a screen device, users inadvertently learn where UI elements are located.
2. *Transfer:* with an appropriate mapping, spatial memory acquired on a physical device can be recalled on an imaginary interface.
3. *Accuracy:* the imaginary interface allows users to point with sufficient accuracy to provide the pointing accuracy required by the associated physical device.

These three assumptions inform the design of transfer-based imaginary interfaces and, in particular, the Imaginary Phone. Next, we support these assumptions with a design discussion and then provide empirical results from two studies.

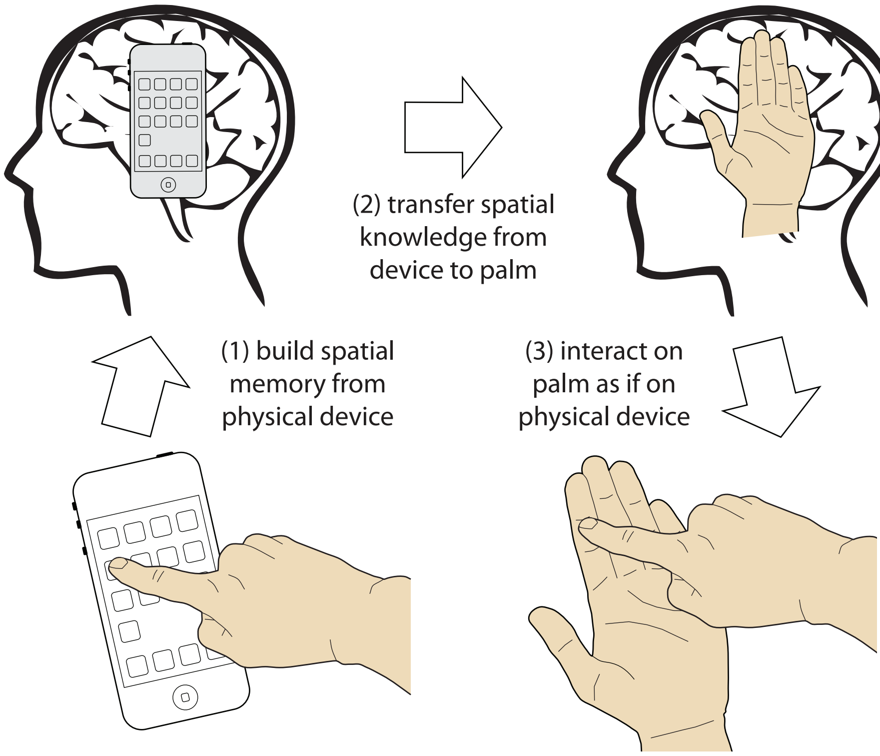


Figure 5.4: The design of Imaginary Phone is based on three assumptions: (1) using a physical device builds spatial memory, (2) the spatial memory transfers to the imaginary interface and (3) users can operate the imaginary interface with the accuracy required by the physical device.

5.2 DESIGN DISCUSSION

Successfully enabling transfer learning depends heavily on the design of the imaginary interface. In particular the last two assumptions (transfer and accuracy) are related to how the system is designed. Here I present the design alternatives we explored and explain the rationale for our choices.

First of all, we chose to use a palm-based imaginary interface. As I discussed in Chapter 4 the palm has much higher selection accuracy than a mid-air imaginary interface. Furthermore, the palm offers a large amount of visual and tactile cues that can be used to associate functions from the physical device to specific locations on the palm.

However, the palm and the device screen generally do not have the same size and shape. This require us to define a *mapping* between screen and palm that is understandable to the user and encourages the transfer of spatial knowledge.

Our prototype uses the simple regular grid mapping illustrated in Figure 5.5b. This layout allows users to imagine the bounding rectangle

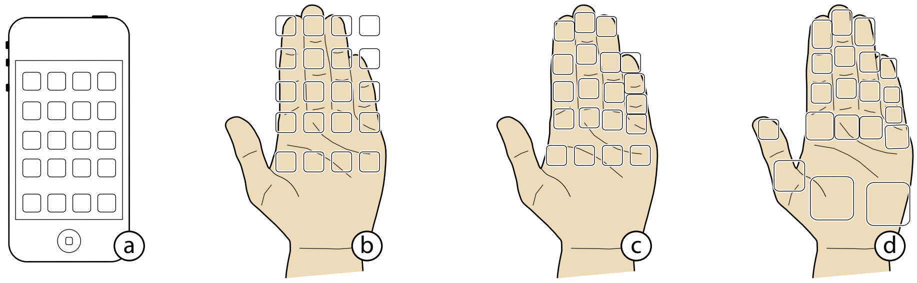


Figure 5.5: Mapping (a) the iPhone home screen mapped to (b) a regular grid, (c) a semi-regular grid where the columns are mapped to fingers and (d) arbitrary mapping of common function to the best landmarks.

of their hand and use that to find the position. Even more importantly, the layout is generic, thus applies to any interface including free form input, such as sketching and handwriting.

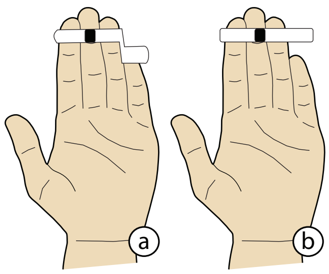


Figure 5.6: (a) Non-regular mappings fail when placing sliders and list items that span the width of the screen. (b) The regular grid works fine.

The more specific layouts (Figure 5.5c and d) should allow for increased pointing accuracy by making even better use of landmarks, but could cause confusion when trying to operate controls that assume a rectilinear screen, such as the slider in Figure 5.6. Highly specialized layouts (Figure 5.5d) are impractical, as they require users to relearn mappings on a per-application basis.

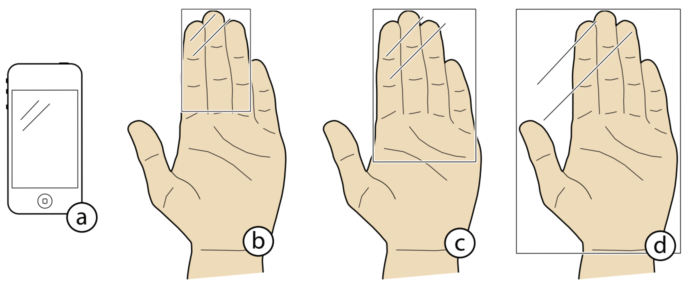


Figure 5.7: (a) The screen on a current 5×7 cm mobile devices (b) maps to approximately three fingers of an adult male hand. (c) Using a scaled mapping allows us to map to four finger or (d) the whole hand. The iPhone and hands are shown in the same scale.

Our Imaginary Phone prototype uses the *4-finger scale* as shown in Figure 5.7c. This maps input from the larger palm to a smaller screen size, resulting in a scale factor of approximately 1.86. Unlike the more obvious 1:1 mapping (Figure 5.7b) the scaling allows us to include additional landmarks and thus increase the effective pointing accuracy on the mobile device. We could continue this logic by increasing the scale to include the whole hand but at the expense of leaving a large amount of the interaction space off the surface of the hand—a space devoid of landmarks.

The four-finger scale works best with interfaces laid out in four column layouts, such as the iPhone home screen (Figure 5.8a). Other layouts, such as a seven-column month calendar (Figure 5.8b) could be mapped by assigning every other day to the space between two fingers, which also make good landmarks. Similarly, we could map three-column layouts (Figure 5.8c) to only the spaces between fingers.

Figure 5.8: iPhone screens laid out in (a) four, (b) seven, (c) and three column grids.

Some mappings, such as the semi-regular grid (Figure 5.5c) are simple enough to communicate with a diagram but with others, such as the regular grid (Figure 5.5b), it is not clear how UI elements map to specific features on the user's hand. Figure 5.9 shows an approach we explored to teach users the regular grid mapping. If we have access to the device's wallpaper, we could display a photo of the user's hand as wallpaper. During use users now learn not only target locations but also the mapping from a widget to its location on the user's palm.

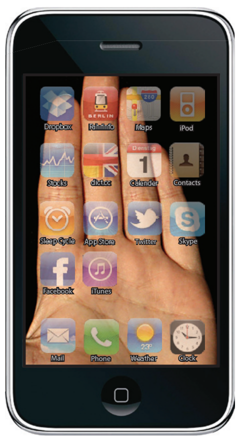


Figure 5.9: A photo of the user's hand as wallpaper helps learn the association between widget and location on the user's palm.

In the following two studies we directly tackle the three assumptions of transfer learning by gathering some empirical results.

5.3 STUDY 1: RECALL AND TRANSFER THE LAYOUT

This first study investigated the first two of the three assumptions behind transfer learning. First, we wanted to confirm that users build up spatial knowledge of an interface through the regular use of a touch-screen mobile device. Second, we wanted to confirm that this knowledge transfers correctly to an interface on the palm. Together, these numbers would tell us *how useful* an Imaginary Phone could be. We did not test a specific quantitative hypothesis in this study; we were purely interested in participants' recall abilities.

During the study we asked daily iPhone users to recall the locations of the (up to) twenty home screen app icons of their own iPhone from memory and without feedback. As a between-subjects design, half of the participants recalled and communicated their choice by pointing to a non-functional iPhone prop (PHONE PROP condition, Figure 5.10a) while the other half recalled locations by pointing to the palm of their own non-dominant hand, using a predefined scheme of how buttons on their iPhone would map to locations on their palm (PALM condition, Figure 5.10b).

5.3.1 Research questions

The goal of this study was to determine how many app locations the participant learned as a side effect of regular use of their mobile device and how much of that knowledge would successfully transfer, using the supplied mapping, to the participants' palm. Participants in the PALM condition would not only have to recall, but also map locations onto their hand. The difference between the two conditions would serve as an indication for how much information is lost in transfer. We also expected the frequency of use to correlate with the user's ability to recall.

5.3.2 Task and procedure

After we seated participants, they unlocked their phone and, without looking at the screen, handed it over to us. Participants in the PALM condition were now taught the semi-regular-grid mapping scheme (see Figure 5.5c in the previous section). This preparation took less than a minute for all participants. See Appendix B.2 for the materials provided to the participant during this preparation.

The experimenter then conducted a series of trials, one for each app on the participant's home screen. For each trial, the experimenter picked a different app randomly and cued the participant with the app's name and a description of the app icon's visual appearance.

Participants responded by pointing to the app's presumed location within the 4×5 icon home screen. Participants in the PHONE PROP condition pointed to cells displayed on a printed prop of an iPhone (the unlabeled all-white icons in Figure 5.10a). Participants in the PALM condition instead pointed to a location on their own non-dominant hand (Figure 5.10b).

In both conditions, the experimenter determined what location the participants were pointing to by observing them point. While we did not measure pointing accuracy directly (we investigated that in Study 2), the experimenter had no difficulty identifying which targets participants referred to.

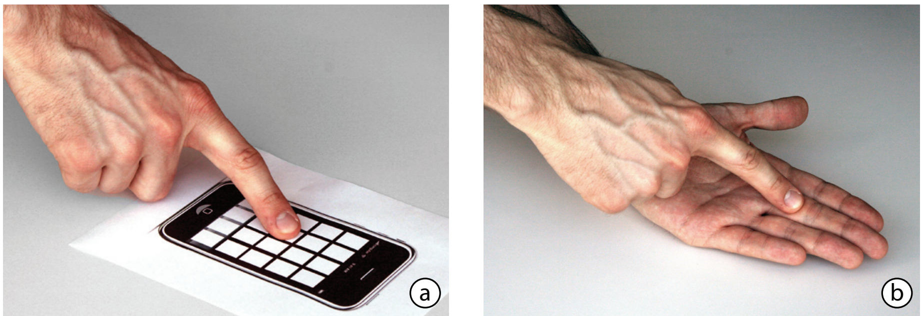


Figure 5.10: Study 1 task – (a) participants in the PHONE PROP condition recalled app locations by pointing to an empty iPhone prop, (b) participants in the PALM condition pointed on their own non-dominant hand.

After completing all trials, participants classified each of their home screen apps as used either *daily* (at least once a day), *weekly* (at least once a week) or *rarely* (less than once per week).

Finally, participants filled out a demographic questionnaire (reproduced in Appendix B.2). All participants completed the study in 15 minutes or less.

5.3.3 Participants

We recruited twelve participants (five female) in the cafeteria of our institution. Participants were on average 23.6 years old (SD = 4.2) and two were left-handed. All participants were daily iPhone users and carried it with them. They were given a small gift for their time.

5.3.4 Results

The twelve participants had on average 18.4 apps on their home screens and recalled each only once, for a total of 221 app recall trials. No outliers were removed but three trials were discarded because of errors by the experimenter, leaving 218 for analysis. Figure 5.11 shows the responses from all participants.

Our main finding was that participants, on average, correctly positioned 64% of the apps on their phone (68% for PHONE PROP, 61% for PALM). The success rate was higher for apps used daily (71% for PHONE PROP, 80% for PALM). In both cases, t-tests did not show any significant differences. When they were wrong, 45% of guesses were only a single cell off, suggesting that participants had some spatial knowledge. Figure 5.12 shows these aggregated results.

Overall the frequency of use of an application correlated with error rate, where less often used applications had a higher error rate (Pearson's $r_3 = 0.998$, $p = 0.043$) but we found no trends relating performance to age, gender or duration of phone ownership.

5.3.5 Discussion

Since none of the participants were aware of the task or project before the study, a mean recall rate of 64% of their home screen apps can only be explained as a side effect of regular phone use. This supports the first of our three main assumptions behind the transfer learning approach: just by using a phone, participants acquire a spatial understanding of the interface.

Also, the observed recall rates from these untrained participants could be considered a lower bound of performance. Actual users of an Imaginary Phone would have an incentive to actively learn locations that would likely result in much better performance.

We did not find a significant difference between recall in the PHONE PROP and PALM conditions. However, while the lack of statistical significance was expected given the small number of participants and high variation, the fact that both numbers are in the same range suggests that the loss of spatial knowledge during transfer cannot be too large. This is also supported by our observations—participants seemed to recall on their hands almost as easily as on a phone. This supports the second of our three main assumptions: spatial knowledge can indeed transfer to the hand.

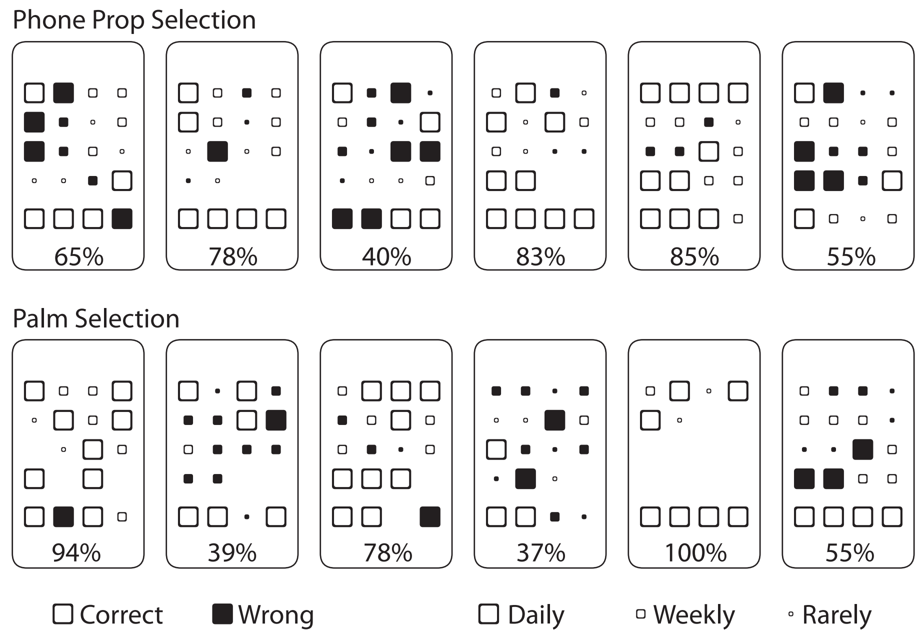


Figure 5.11: Study 1 raw results – each of the twelve large rounded rectangles represents one participant's phone home screen. Each black (wrong) or white (correct) square represents one home screen app. Percentages indicate each participant's recall rate.

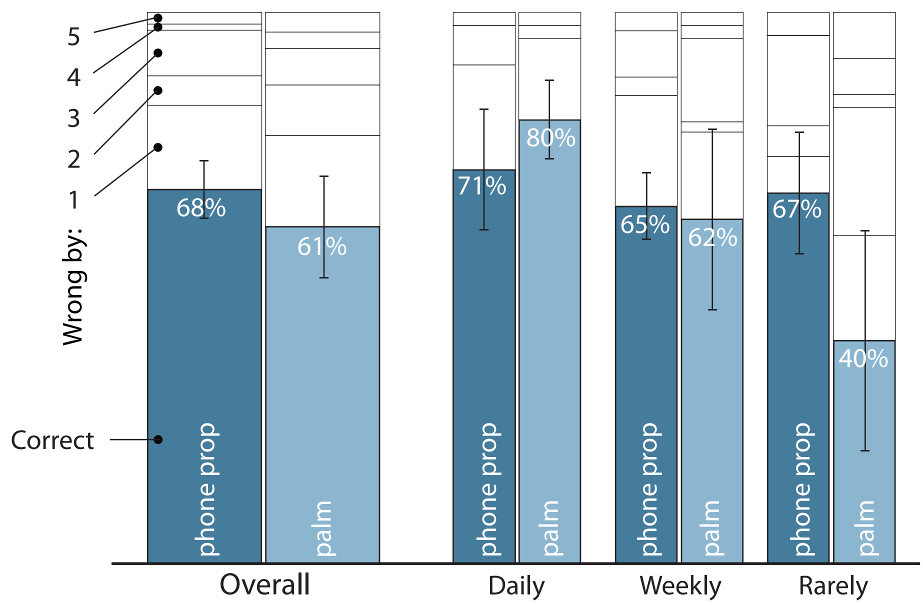


Figure 5.12: Study 1 aggregated results – percentage correct by use frequency (± one standard error of the mean). The chart is stacked with mean percentages for incorrect responses separated by how far (in Manhattan distance) they were wrong by.

5.4 STUDY 2: RAW TARGETING PERFORMANCE

Our goal of this study was to verify the third main assumption required for the Imaginary Phone interaction style: that targeting on the palm can be done with sufficient pointing accuracy to operate the typical functions of a mobile phone. In particular, we wanted to know whether this interaction style would allow users to select the widget sizes common to today's touch devices. If so, imaginary interfaces that mimic touch devices would become viable.

Therefore in this study we measured the raw selection performance of the 20 targets that represent the application launch icons on the home screen of the iPhone. To verify the experiment's validity we also included selection on an iPhone as a control condition. At the same time we were interested in how well users would be able to use the Imaginary Phone in an eyes-free manner, and we included a blindfolded condition to study this.

5.4.1 Research questions and variables

The main research question we set out to answer with this study was "what is the smallest size buttons that participants would be able to select on an Imaginary Phone, using their palm as an interaction surface". If the minimum button size was smaller or similar to the size of UI elements in the iOS standard widget library then users of an Imaginary Phone should be able to mechanically operate any iOS application running on an iPhone.

We also used this as an opportunity to confirm if palm-based imaginary interfaces could be used in an eyes-free manner. Consequently we included a blindfolded condition in the study.

The study used a 2×2 within-subjects factorial design. The two factors were:

INTERACTION SURFACE: PHONE VS. PALM

SIGHTEDNESS: SIGHTED VS. BLINDFOLDED

5.4.2 Apparatus

To provide high tracking accuracy we used an fixed optical motion capture system (OptiTrack with eight V100 cameras) that tracked retro-reflective markers. For the PALM condition we attached a marker set to back of the participants' left hand and a separate market set on the pointing finger of their dominant hand (as shown in Figure 5.13).

The reflective marker set placed on the participants' pointing finger was attached such that one of the markers was directly above the participants' index finger nail. We configured the motion capture software to report the position of this finger nail marker (by default the center of the three marker set would be reported). This provided a stable and rotation invariant position that we used to calculate the on-hand touch position.

Figure 5.13: Study 2 apparatus – (a,b) participants stood in the center of an optical tracking environment with markers attached to the back of their non-dominant hand (to track the interaction surface) and the index finger of the dominant hand (to track where they were pointing). (c) A closeup of the marker set over the participant's index finger.

Each participant calibrated the system with a 23-point calibration procedure: three initial points are used to calculate the plane of the hand and the remaining 20 points isolate the precise location of each imaginary button on that plane. To control for differences in the sensing mechanism, the PHONE interface was implemented in the same manner with a marker set attached to the top of a phone and the selection point determined by where the marker set attached to the participants' pointing finger was located with respect to the interaction surface of the phone.

5.4.3 Task and procedure

During the study, participants repeatedly selected targets on their palm as prompted by the system. Participants started each trial by pressing a footswitch, which signaled to the system to verbally announce and display a description of a grid location that corresponded to the grid location on the interface (using the semi-regular grid described in Section 5.2). For example, the system would announce and display "col

3, row 1" to indicate that the participant should select the tip of the ring finger (third finger, first segment). At the same time, the system displayed an image depicting the correct selection location overlaid onto an image of a hand or a phone, depending on the condition. Participants placed their pointing finger at the location they believe corresponded to the prompted location and confirmed the selection by pressing the footswitch a second time, completing the trial.

Since participants' task was to select targets at the position that they chose in the calibration procedure, we were able to control for differences in hand sizes and the minor differences in how participants thought the targets should be arranged on their hand. For instance, some participants were inclined to place targets directly at the end of their finger tips, while others thought the center of the last finger segment was more natural. By allowing the participants to choose the location themselves, we avoided any confusion on the participants' part.

In each block, participants selected each of the 20 grid locations two times each as they were presented in random order. There were four blocks per participant, each using one combination of the independent variables, presented in counterbalanced order based on a balanced Latin square. After performing all blocks, participants completed a short questionnaire (reproduced in Appendix B.3). All participants completed the study in 30 minutes or less.

5.4.4 Participants

We recruited a separate set of 12 participants (three female) from our institution. They were 21 to 29 years old (M = 24.5, SD = 2.7) and all but one were right handed. Each was given a small gift in exchange for their time.

5.4.5 Analysis approach

With the Imaginary Phone, the user's touch positions on their palm are transferred to touches on their phone. Therefore to properly validate this interaction and to allow comparison to a phone-based interfaces, we normalized the palm selection positions to match how they would appear on the phone. This was accomplished by scaling the interaction surface of the hand down to the phone surface, while maintaining the irregular position of the targets on the palm.

From the normalized selection points we followed an approach similar to what we used in Section 3.5 and Section 4.3, which was based on the approach of Holz and Baudisch (2010). We decomposed selection error into *offset* (constant error) and *minimum button size* (spread or variable error). Unlike the previous studies in this dissertation, here we assumed per participant calibration by calculating the offset for each participant and target.

> *Note:* the offsets and minimum buttons sizes reported in the following subsection are the distances on the phone interface after they have been resized down from the palm interface. The raw distribution of selection positions on the palm are approximately 1.86 times larger (the average American male index finger is 7.53 cm and the average female index finger is 6.96 cm (Greiner 1991), which average to 7.25 cm, 1.86 times larger than the equivalent length on the screen interface).

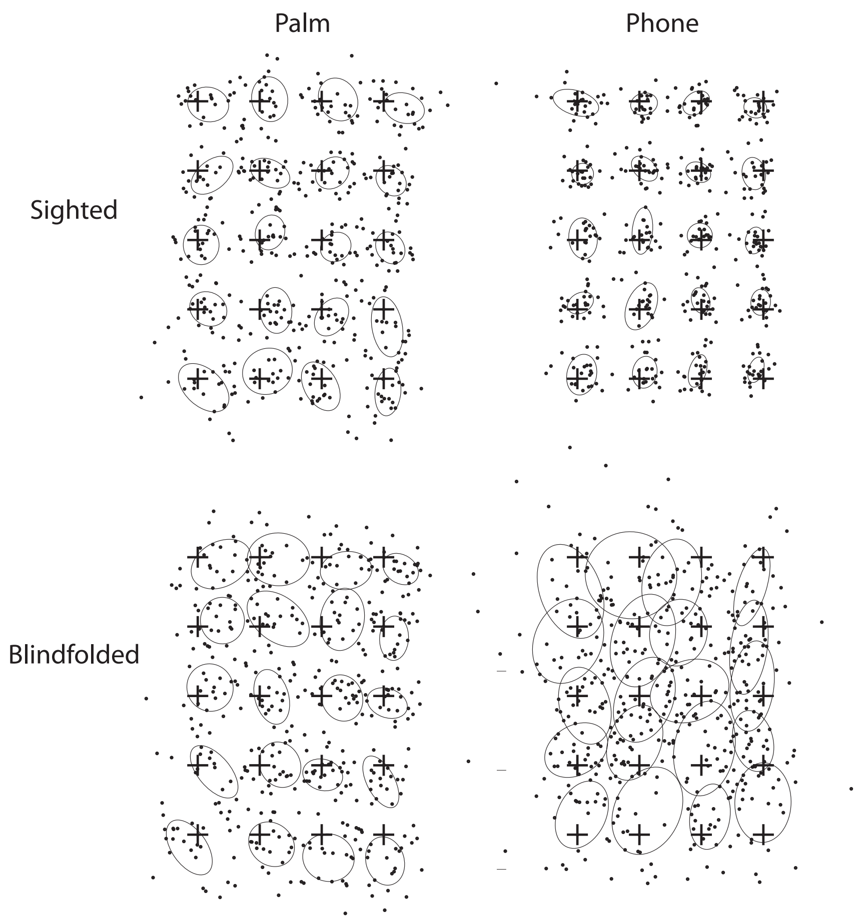


Figure 5.14: Study 2 raw results – the cross represents the target location and each dot is the selection point from one trial. The ovals provide an indication of the minimum button sizes by encompassing 68% of the selections per target and participant (the extent of two standard deviations of the binomial distribution derived from responses).

5.4.6 Results

We gathered data from 1920 trials and after removing 5 outliers, 1915 trials remained in the analysis. Outliers were defined as being more than three standard deviations away from the overall mean for the condition.

The raw data collected during the experiment are shown in Figure 5.14. Each point is one selection from a study participant. They are grouped by ovals that provide a represention the minimum buttons size for each target per participant.

Offset

As shown in Figure 5.15, the average error offset for SIGHTED trials on the PALM was 1.93 mm (SD = 1.96) to the right and 0.86 mm (SD = 1.85) down. On the PHONE the offset was 0.15 mm (SD = 1.42) to the left and 0.50 mm (SD = 1.82) up.

When blindfolded, the offsets were relatively worse with the PALM registering 2.45 mm (SD = 2.25) to the right and 1.42 mm (SD = 1.95) down, while trials using the PHONE had an average offset of 0.97 mm (SD = 3.1) to the left and 0.12 mm (SD = 4.30) up. Note the relatively high standard deviations from blindfolded phone condition; upon inspection of Figure 5.14 it becomes clear why the average offset disappears in the BLINDFOLDED PHONE condition—the whole interaction area was compressed vertically with both the top and bottom rows trending towards the middle.

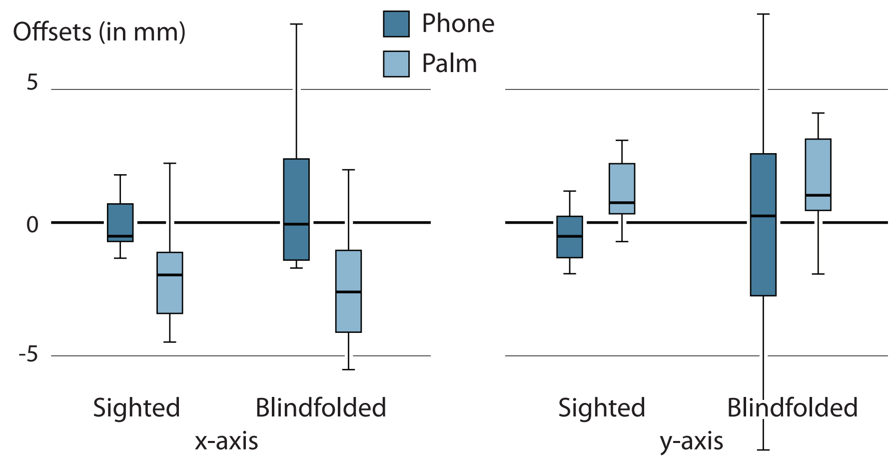


Figure 5.15: Study 2 aggregate offsets – boxplots showsing the distribution of offset distance (separated by axis) from target to center of responses.

A repeated measures ANOVA showed that offsets on the PALM were significantly different than on the PHONE along the x-axis ($F_{1,11} = 16.252$, $p = 0.002$, $\eta^2 = 0.285$) but not along the y-axis. The ANOVA also found a significant interaction effect between the two conditions along the x-axis ($F_{1,11} = 5.67$, $p = 0.036$, $\eta^2 = 0.0177$). Follow-up t-tests showed that the PHONE and PALM resulted in significantly different x-axis offsets when SIGHTED ($t_{11} = 3.298$, $p = 0.007$, Cohen's $d = 1.989$) and BLINDFOLDED ($t_{11} = 4.094$, $p = 0.002$, Cohen's $d = 2.469$).

While the offset values are interesting to get a feeling for how well participants perform the interaction, they can be easily corrected in software with a simple function that is used to subtract the preobserved offset from each touch position. Minimum buttons sizes are more useful to determine how well users of a system will be able to target locations in day-to-day operation.

Minimum button size

The aggregate minimum buttons sizes for each condition are shown in Figure 5.16. To determine the statistical significance of these results we ran repeated measures ANOVAs for minimum button sizes in x and y. For both axes there was no significant main effect of INTERACTION SURFACE. Blindfolding, on the other hand, did have a significant effect along the x-axis ($F_{1,11} = 94.109$, $p < 0.001$, $\eta^2 = 0.432$) and the y-axis ($F_{1,11} = 78.99$, $p < 0.001$, $\eta^2 = 0.482$). There was also an interaction effect between the two variables along the x-axis ($F_{1,11} = 15.805$, $p = 0.002$, $\eta^2 = 0.120$) and the y-axis ($F_{1,11} = 45.882$, $p < 0.001$, $\eta^2 = 0.203$).

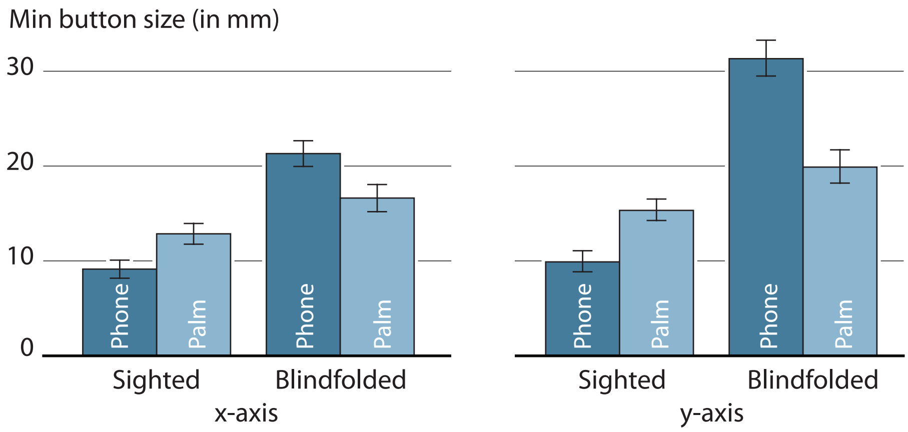


Figure 5.16: Study 2 aggregate minimum buttons size – average size (separated by axis) of the buttons that would capture 95% of the participants' selections. Error bars are ± one standard error of the mean.

When sighted, the average minimum button size for the PALM condition was 12.86 mm (SD = 3.77) along the x-axis and 15.34 mm (SD = 3.92) along the y-axis. In the PHONE condition the average minimum button size was 9.13 mm (SD = 3.30) along the x-axis (28.9% less than the palm) and 9.91 mm (SD = 3.85) along the y-axis (35.4% less). Follow-up t-tests did not show a significant difference between the button sizes on the x-axis ($p = 0.031$, above the Bonferroni corrected α of 0.025), however on the y-axis the difference was significant ($t_{11} = 3.807$, $p = 0.003$, Cohen's $d = 2.296$).

Blindfolding the participant reverses the trend; the PALM becomes the more accurate surface. In this case, the average minimum button size for PALM is 16.63 mm (SD = 4.95) in x and 19.90 mm (SD = 6.07) in y, whereas the button size dimension for the PHONE condition are 21.31 mm (SD = 4.70) and 31.33 mm (SD = 6.56), which are 28.2% and 57.5% larger respectively. Again, these differences were not significantly different on the x-axis ($p = 0.035$) but they were on the y-axis ($t_{11} = 4.127$, $p = 0.002$, Cohen's $d = 2.489$).

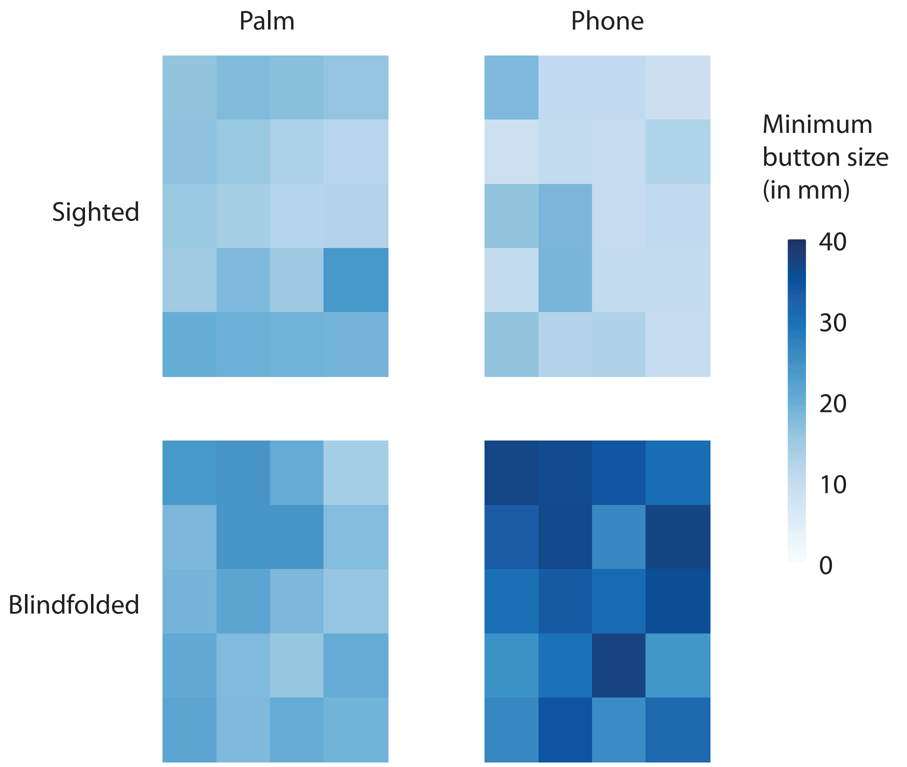


Figure 5.17: Study 2 minimum circular button size for each target locations (represented as colored rectangles in a grid) for the four experiment conditions. The darker the cell the larger the button.

Finally, to check if participants performed differently for different target locations we also looked at the minimum button size for each target location. The results are shown as a heatmap separated out by target in Figure 5.17. In this figure, we simplified the results by combining variable error along the x and y dimensions into one metric by taking the maximum of the two, effectively computing the minimum *circular* button diameter. Although there is considerable variation among the target locations, there is no clear pattern. Perhaps there is a small effect along the bottom row for the palm when sighted, which would correspond to the less defined space on the fleshy part of the palm as compared to the finger segments. There may also be a slight negative effect along the left column for the phone when sighted, perhaps because of interference with the thumb of the supporting hand. However, we must be very careful deriving insights from these patterns as they may also simply be noise a result of inconsistent targeting by the participants.

From the questionnaire, 9 of the 12 participants preferred the phone as an interaction surface over the palm when sighted (the other two preferred the palm and one was undecided). However, when blindfolded the preferences matched their performance with 11 of 12 preferring the palm as the interaction surface.

5.4.7 Discussion

The official human interface guidelines for iOS, which governs the design of applications for the iPhone (Apple Inc. 2012), states that the minimum size for a tappable UI element is 44×44 points (or 15.52×15.52 mm). Our study yielded an average minimum button size of 12.86×15.32 mm indicating that a palm-based imaginary interface provides the necessary accuracy to acquire standard widgets on current touch devices, such as the iPhone, therefore confirms our third assumption about transfer learning.

The observed offset error was also reasonably small, but since this can be easily corrected it does not affect the participants ability to operate the Imaginary Phone interface.

Overall, the raw selection accuracy obtained in this study with both the palm-based imaginary interface and the phone-based control condition are in the same order of magnitude as the accuracy values obtained by other researchers with modern touchscreens. They report, using different study conditions (that make the results difficult to compare directly), minimum button sizes of 15.0 mm (Holz and Baudisch 2010), 11.5 mm (Wang and Ren 2009) and 10.5 mm (Vogel and Baudisch 2007).

The study from the previous chapter (in Section 4.3), that compared selection of target in the mid-air to targets on the palm, found that palm-based target had a circular minimum button diameter of 17.7 mm. However, to allow comparison we must scale our current findings back up to the palm (by multiplying by 1.86). By doing that we obtain an on-palm circular button diameter of 28.53 mm, which is substantially larger than the previous result.

The result from Section 4.3 differs for several reasons. First, the previous study used a different sensing mechanism: touch locations were determine post-hoc by the experimenter from the analysis of high-resolution photographs, providing an extreme level of accuracy that the current study apparatus could not replicate. Second, the participants' task in the previous study was simpler with only three targets they needed to repeatedly select, compared to 20 in the current study.

With its superior (but highly impractical) sensing mechanism and simpler task, the results from the previous study can be considered a theoretical maximum of selection accuracy. Whereas the results in this study more closely represent what might be possible with practical sensing processes.

In the present study we also compared sighted to blindfolded use. We found that blindfolding does significantly affect accuracy overall and that the phone interface was especially affected. It appears that the surface of the hand offers extra cues that allow it to perform relatively better when blindfolded compared to the featureless surface of the phone. This finding has large potential implications that I will return to in Chapter 6 with an in-depth investigation.

5.5 SUMMARY AND DISCUSSION

Summarizing the two studies reported above, we found support for all three of the assumptions of transfer learning stated earlier. We found that:

1. Users indeed build up spatial memory automatically while using a physical device: participating iPhone users knew the correct location of 68% of their own iPhone home screen apps without training.
2. Spatial memory can be transferred from a physical to an imaginary interface: participants recalled the location of home screen apps with 61% accuracy when pointing on the palm of their hand.
3. Raw selection on the palm is precise enough to allow operating the device: participants could reliably acquire 12.86×15.32 mm targets on the phone by selecting on their palm. This is sufficient to operate standard widgets on today's mobile touch devices.

Combined these results provide strong support that the transfer learning concept is viable from a human performance point of view, which along with our Imaginary Phone prototype providing the proof of concept implementation, shows the potential a transfer learning based imaginary interface.

CONCLUSIONS

In this chapter, I presented a method of learning imaginary interfaces based on transfer learning and illustrated the concept with our Imaginary Phone prototype.

From the perspective of a mobile device user, the main benefit of imaginary interfaces based on transfer learning is that they allow operation of a mobile device without actually retrieving it from the pocket or purse. What is promising is that a similar transition has happened before. As illustrated in Figure 5.18, early devices required users to retrieve device and stylus (e.g., the PalmPilot) and eventually usage transitioned to touchscreen-based devices that did not require a stylus. This move took place even though stylus input is in many ways superior to touch input—it offers higher precision (no fat finger problem (Vogel and Baudisch 2007; Holz and Baudisch 2010)).

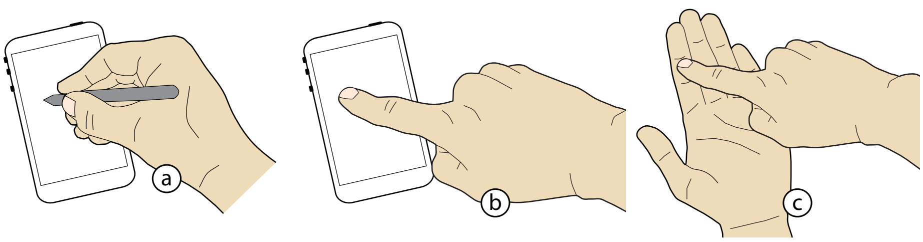


Figure 5.18: (a) Early mobile devices required users to retrieve a stylus and the device. (b) Current touch devices require retrieving only the device. (c) Imaginary interfaces do not require retrieving anything.

At the expense of losing even more precision and essentially limiting interaction to microinteractions, systems like the Imaginary Phone have the potential to offer even more convenience: there is no longer any need to retrieve the device itself. Even if, at this point, imaginary interfaces hardly viable on their own, the combination of an imaginary interface with a physical mobile device has as potential as a form factor for tomorrow's mobile interfaces.

To provide a solid foundation for Imaginary Interfaces, the next chapter looks at Imaginary Interfaces from another perspective. I look beyond the design and engineering aspects of Imaginary Interfaces and investigate not what is possible with an imaginary interface but what human perceptual abilities allow them to function.

6 UNDERSTANDING IMAGINARY INTERFACES

In this chapter I present the results of an exploration to provide a deeper understanding of Imaginary Interfaces, i.e., not what they allow users to do, but *why* they allow doing it. In the bulk of this dissertation, I presented concepts and studies that showed what is possible with Imaginary Interfaces but in this chapter I present a much lower level exploration into what human perceptual abilities enable this interaction style.

To do so, we adapted an interface for visually impaired users to a palm-based imaginary interface, shown in Figure 6.1, and ran a series of lab studies to explore which inherent properties of palm-based imaginary interfaces cause them to perform as well as they do.

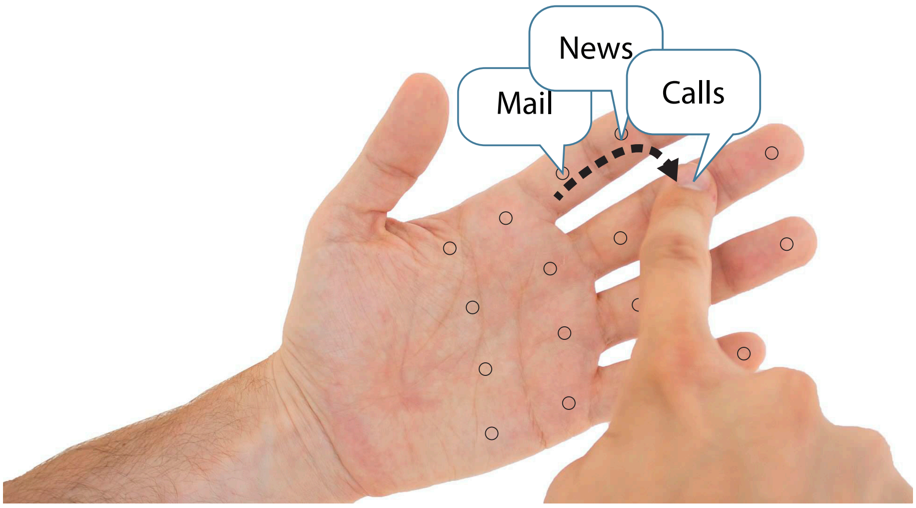


Figure 6.1: We adapted a non-visual audio interface that announced targets as users touch them. This allows users to browse an unfamiliar imaginary interface, forming the basis for our exploration.

Contributions

The main contribution of this chapter is an exploration into the inherent properties of palm-based imaginary interfaces and how these properties are responsible for user performance. We found that (1) visual cues, i.e., observing ones hands performing the interaction; (2) tactile cues sensed by the palm and (3) tactile cues sensed by the pointing finger all contribute to performance, in that order.

These findings deepen our understanding of Imaginary Interfaces and confirm that palm-based imaginary interfaces enable stand-alone eyes-free use. Since it was the extra tactile cues available on the palm that improved interaction, these findings motivate future work on on-body eyes-free interfaces, especially for visually impaired users. We investigate this final implication with an exploratory study and interview with one blind participant that confirms our findings.

6.1 INTERACTION STYLE FOR THE STUDIES

Inspired by touch-and-explore interfaces designed for visually impaired users (Talking Fingertip Technique (Vanderheiden 1996), SlideRule (Kane et al. 2008) and most recently the commercial VoiceOver for iPhone) we created an palm-based audio feedback interface that announces targets as users scrub across them.

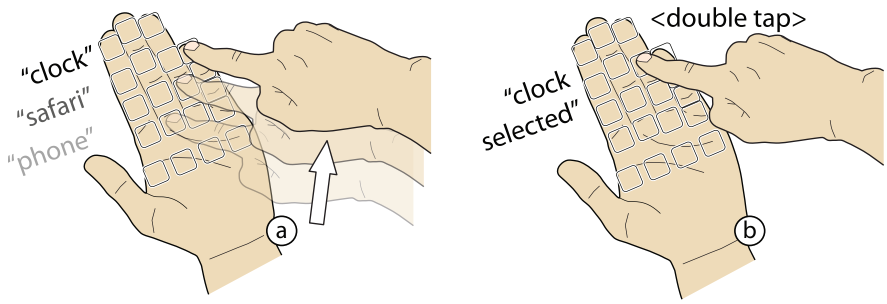


Figure 6.2: We adapted the touch-and-explore style interaction to imaginary interfaces: (a) as users scrub along their palm, the system announces the name of the function at each location. When users find what they are looking for (b) they double tap to perform the selection.

We adapted this interaction style for use with imaginary interfaces. Figure 6.2 shows the resulting interface based on the Imaginary Phone (see Section 5.1). As users drag their fingers across the palm surface, they enter different buttons and the system responds by announcing the name of the target, such as "clock". If users continue further, the auditory feedback is immediately interrupted and the new name is announced. This allows users to quickly scan through lists of items by only listening to the first sound in each word. Users familiar with the layout can also shortcut this exploration and acquire a target by tapping directly on it. We refer to this exploratory action using the supplied audio feedback as *browsing* an imaginary interface.

6.2 OVERVIEW OF STUDIES

As seen in the previous two chapters, the palm-based imaginary interface performed very well in the scenarios we tested. Encouraged by this, we conducted three user studies with the goal to determine which properties of this style of imaginary interfaces were responsible for their performance.

While Imaginary Interfaces share properties with interfaces for visually impaired users—neither relies on visual feedback—they offer extra cues that are potentially relevant:

1. *Visual cues:* While the lack of a screen prevents imaginary interfaces from providing actual dynamic feedback, they do offer the ability for the users to watch themselves interact, providing an extra feedback channel.
2. *Tactile cues:* During interaction with a palm-based imaginary interface, users' hands touch, which provides them with tactile cues in two directions: the pointing finger feels the palm and the palm feels the pointing finger.

To explore the role of these cues we ran three user studies. Each study had the same general form but differed by the surface used for interaction. Figure 6.3 shows the six different interface conditions that participants used throughout the three studies and how they relate to each other.

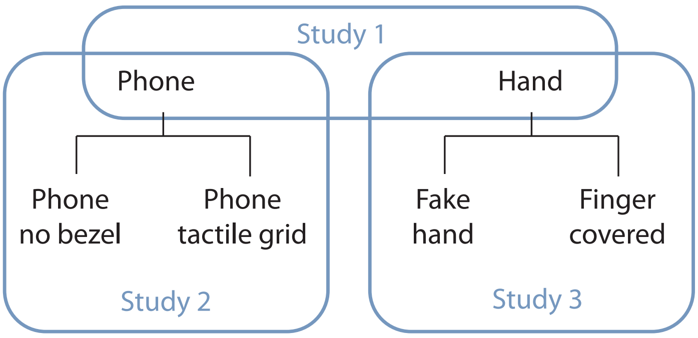


Figure 6.3: The form factors used in the three user studies in this chapter.

Study 1: visual cues

In Section 5.4 we showed that raw selection performance depends strongly on visual feedback (especially on a phone surface), however the dynamic audio feedback helped to direct the user's interaction, preventing us from discovering if visual cues were necessary for this interaction. In this study we explicitly asked this question: Does watching your hands support interaction with a browsing-style imaginary interface? We explored this by comparing blindfolded with sighted use (neither with visual screen feedback) on the phone and the palm.

The results of Study 1 showed that watching your hands interact improves performance and we received first insights about the role of tactile cues: blindfolded interaction did better on the palm than on the phone. However, it remained unclear if the extra tactile cues on the palm were responsible. To explore this we ran two more studies: the first focused on the tactile sensation in the pointing finger, the second on the tactile sensation on the palm.

Study 2: tactile cues sensed by the pointing finger

In this study, we compared three versions of the phone interface, which the participants operated while blindfolded. The first was a plain touchscreen phone and the second was the same with an engraved tactile grid. In addition, we were wondering whether the phone was really featureless or whether touching the bezel and the supporting hand helped users to orient themselves. To investigate this we added a third condition that embedded the phone interaction surface into a large clear piece of acrylic, thereby preventing participants from using the bezel to obtain tactile cues. We included sighted use as an additional baseline.

Study 3: tactile cues sensed by the palm

To study the tactile cues sensed by the palm, we created another two interfaces that participants used while blindfolded, in addition to the palm interface from Study 1. We compared interaction on the palm to interaction on a silicone cast of a hand and to interaction on the palm with a covered pointing finger that minimized fine tactile cues sensed by the finger. Again, we included sighted use as an additional baseline.

Apparatus used in all studies

As shown in Figure 6.4, the participant sat in front of a table with a monitor showing instructions located directly in front of them. A footswitch was used to confirm selection.

To provide high tracking accuracy, we implemented the browsing interaction style using an OptiTrack motion capture system with six V100 cameras that tracked reflective markers. As shown in Figure 6.5, the participant's non-dominant hand was placed in a fixture molded to the back of their hand. Using moldable putty, we created an imprint of the back of the participant's hand at the start of each experiment. The fixture allowed the participant to replace their hand in the same position when switching between the PHONE and PALM conditions while maintaining a consistent calibration. To track where the participant is

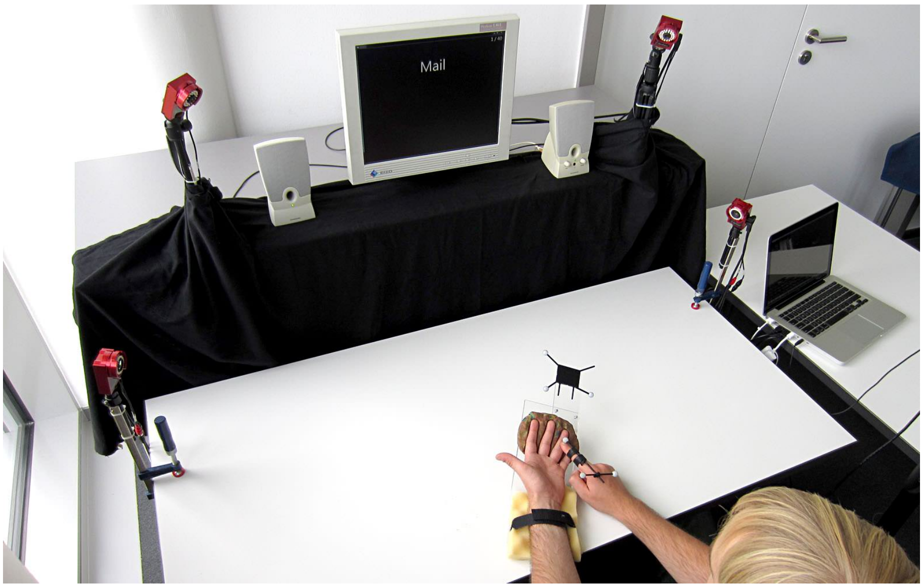


Figure 6.4: Apparatus used in the following studies. The participant's non-dominant hand was fixed to a brace to ensure consistent calibration and the participant's pointing finger was tracked with reflective markers. A footswitch (not shown) was used for confirmation.

touching the palm, we attached a marker set to the index finger of the participant's dominant hand such that is was aligned with the extended index finger.

When the participants' index finger was within 3 mm of the plane defined on their non-dominant palm, the interaction was in the *touching* state that provided invoked the audio-feedback that enabled participants to explore the interface.

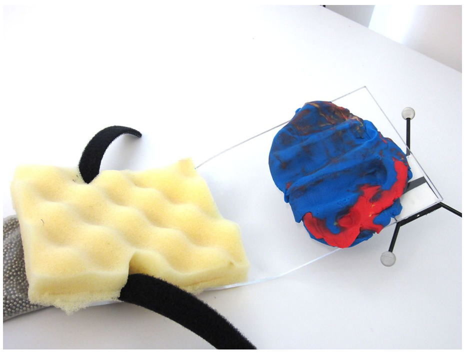


Figure 6.5: Closeup of the hand brace fixture to ensure consistent calibration.

Participants calibrated the system with a 23-point calibration procedure: three points were used to find the plane of the hand and the remaining 20 to find the precise location of each finger segment that were mapped to imaginary button locations. We were careful to leave the participants' palm and pointing finger unobstructed in order to not interfere with the interaction between the two hands.

6.3 STUDY 1: THE IMPACT OF VISUAL CUES

With this study we tested the role of visual cues when using a browsing-style imaginary interface. Specifically, we wished to confirm the conjecture from Chapter 3 that the visuospatial memory created by watching your hands interaction is the feedback channel that enables use of imaginary interfaces.

To that end, here we compared performance while sighted and blindfolded using our browsing interface on the phone and the palm.

6.3.1 Apparatus, task and procedure

The study used a within-subjects 2×2 factorial design with these independent variables (shown in Figure 6.6):

SIGHTEDNESS: SIGHTED VS. BLINDFOLDED

INTERACTION SURFACE: PHONE VS. PALM

For the palm condition, participants used the general apparatus described earlier. For the PHONE condition, we tracked interaction with the same optical tracker system used in the palm condition. This kept any potential tracking errors consistent across conditions. The phone used in the study was a non-functional replica of an iPhone 3G with identical surface area but thinner (at 5.5 mm).

During the BLINDFOLDED conditions, a partial blindfold was used (shown in Figure 6.6b). Due to its shape, it obscured the participants' view of their hands but not of the display in front of them.

In each trial participants searched for and selected a prompted target. They started the trial by pressing a footswitch. The system then spoke the target name and showed it on a screen facing the participant. The participants touched the interaction surface with their finger and as they moved it around, the system announced the name of each target. When participants found the required selection they pressed the footswitch to complete the trial. We measured task time from the start of the trial until the participant made a selection. If the selection was incorrect, the trial was discarded and the participant was required to repeat the trial.

Before beginning the experiment, participants received instructions on how to use the system and performed a series of practice trials with each interaction surface until they indicated they understood the interaction style and were comfortable with the system.

During each of the four blocks (each tested one combination of variables) participants had to repeatedly locate five targets out of the 20

Figure 6.6: Study 1 conditions – (a) SIGHTED vs. (b) BLINDFOLDED, using a partial blindfold that only obscures the participants' view of their hands; (c) PHONE vs. (d) PALM.

available targets in the interface. The five targets (chosen randomly) were presented to the participants eight times in random order.

We presented the conditions in a counter-balanced order using a balanced Latin square. Each condition used a different set of target names derived from a survey of the most popular iPhone apps used by local students (see Appendix A).

At the end of the experiment participants completed a short questionnaire to gather their preference of interaction surface when blindfolded and not (reproduced in Appendix B.4).

6.3.2 Hypotheses

Since vision and taction work together (Làdavas et al. 2000), we expect that the sighted conditions (that combine both visual and tactile cues) will outperform the conditions where only taction is available. Therefore, in our first hypothesis we wish to confirm this idea:

HYPOTHESIS H1 Participants will be faster when SIGHTED.

However, since taction far outperforms proprioception (Fuentes and Bastian 2010), we believe that the tactile cues available on the palm are more likely to be able to fill in for visual cues when they are not present, compared to the mostly featureless phone surface. Therefore our second hypothesis is:

HYPOTHESIS H2 When BLINDFOLDED, using the PALM as an interaction surface will result in faster search times.

6.3.3 Participants

We recruited 12 participants (2 female) from our institution. They ranged in age from 22 to 30 (M = 26.0, SD = 2.63). All were right-handed and all had normal or corrected to normal vision and hearing.

6.3.4 Results

We collected 1938 data points and removed 18 error trials (0.9%) and 43 outliers (2.2%), leaving 1877 trials in this analysis. We defined outlier response times as three standard deviations above the mean for each condition and repetition. Participants completed the study within 30 minutes.

We ran a $2 \times 2 \times 8$ (SIGHTEDNESS $\times$ INTERACTION SURFACE $\times$ repetition) repeated measures ANOVA on completion time. There was no overall significant difference between PHONE and PALM ($p = 0.11$) but when BLINDFOLDED participants were 50% slower than when SIGHTED (5.39 sec. vs. 3.59 sec., $F_{1,11} = 99.90$, $p < 0.001$, $\eta^2 = 0.08$), which confirms our first hypothesis that watching your hands improves interaction.

As shown in Figure 6.7, there is a clear learning effect ($F_{1,11} = 85.55$, $p < 0.001$, $\eta^2 = 0.54$) and by inspection one can see that the participants' selection times steadily decrease in the first three or four repetitions then level off in the remaining repetitions.

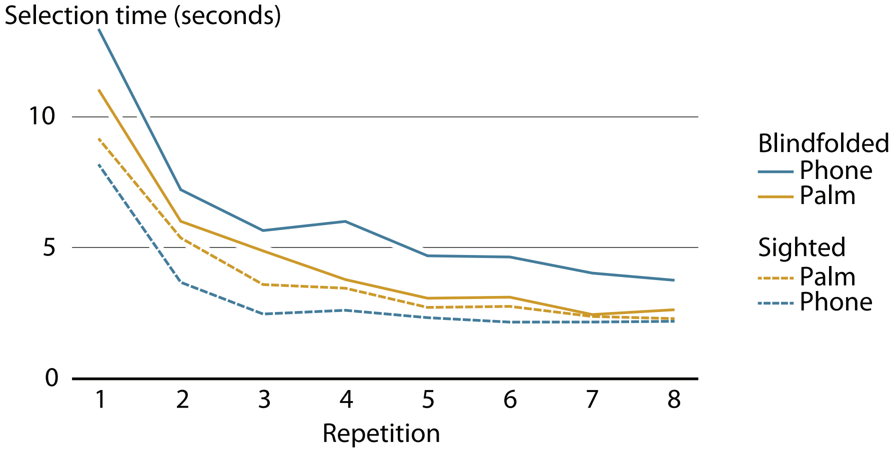


Figure 6.7: Study 1 results showing performance over time.

To investigate these results further we aggregated the repetitions into two equal blocks: *learning phase* (the first four repetitions where participants acquired knowledge of the target locations, results shown in Figure 6.8a) and *trained phase* (the last four repetitions where participants had acquired good knowledge of the target locations and response time have leveled off, results shown in Figure 6.8b). We then analyzed each with a separate 2×2 repeated measures ANOVA.

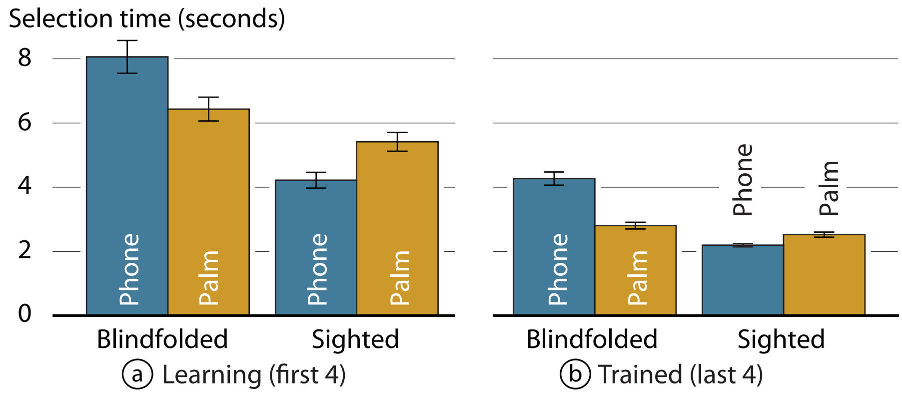


Figure 6.8: Study 1 aggregate selection times for the (a) first and (b) last half of the repetitions. Error bars are ± one standard error of the mean.

Learning phase (Figure 6.8a)

In this phase participants took approximately 6 seconds on average to find and select each target with only slight differences between the PHONE (6.15 sec.) and PALM (5.92 sec.) condition.

When BLINDFOLDED, participants were 50% slower than when SIGHTED (7.25 sec. vs. 4.82 sec., $F_{1,11} = 66.26$, $p < 0.001$, $\eta^2 = 0.41$) and there was an interaction effect between SIGHTEDNESS and INTERACTION SURFACE ($F_{1,11} = 9.72$, $p = 0.01$, $\eta^2 = 0.14$).

Looking at BLINDFOLDED and SIGHTED trials separately, we see that when using the PHONE being BLINDFOLDED resulted in a 91% worse task time (4.22 sec. vs. 8.06 sec.; $t_{11} = 8.84$, $p < 0.001$, Cohen's $d = 5.33$) but when using the PALM being BLINDFOLDED only led to 19% worse performance (5.41 sec. vs. 6.43 sec.). This last difference was not significant ($p = 0.14$).

Trained phase (Figure 6.8b)

In this phase, participants were overall 18% faster to select a target on the PALM compared to the PHONE (2.66 sec. vs. 3.23 sec.; $F_{1,11} = 15.33$, $p = 0.002$, $\eta^2 = 0.07$).

When BLINDFOLDED participants were 50% slower than when SIGHTED (3.53 sec. vs. 2.36 sec.), also a significant difference ($F_{1,11} = 66.54$, $p < 0.001$, $\eta^2 = 0.29$). Like in the learning phase there was a significant interaction effect between INTERACTION SURFACE and SIGHTEDNESS ($F_{1,11} = 21.55$, $p = 0.001$, $\eta^2 = 0.159$).

This interaction effect comes from the fact that when using the PHONE, participants who were BLINDFOLDED were 94% worse than when SIGHTED (2.19 sec. vs. 4.27 sec.; $t_{11} = 6.27$, $p < 0.001$, Cohen's $d = 3.78$) but were not significantly worse when using their PALM (2.51 sec. vs. 2.80 sec., $p = 0.136$).

Questionnaire

When BLINDFOLDED, 11 participants preferred to use their PALM (one had no preference) and 10 participants rated the PALM faster than the PHONE with the remaining two rating the PHONE faster. When SIGHTED, the preference was split with five participants for each interface (two had no preference) but eight indicated the PHONE was faster and two that the PALM was faster (two reported neither).

Participants commented that when blindfolded the palm offered more tactile cues and the phone lacked a "reference system". One said, "There are more features on the hand. On the hand you can relate terms to fingers." However, many commented that when not blindfolded the straightforward grid of targets on the phone was easier to traverse: "When not blindfolded the grid helps to be more efficient." One participant noted that the tactile cues were sufficient even when not blindfolded, stating, "Even in 'sighted' mode I'd rarely look at the phone/hand anymore once I learned the positions."

6.3.5 Discussion

In the study we found that participants performed better when they could see their hands interact. That is, despite receiving no visual feedback from the device itself (on the screen for example), the participant's vision was nonetheless necessary to record and recall the spatial locations of the targets on both the phone-based interfaces and the palm-based interface. This confirms our assertion in Chapter 3 that the user's visuospatial memory is what provides the necessary feedback to enable interacting with an imaginary interface.

Furthermore we gathered first insights into how tactile cues on the palm contribute to eyes-free use. However, we only know that the palm performed better when blindfolded. This was presumably because of the extra tactile cues available on the palm, especially when compared to the mostly featureless phone surface. However, we do not know which tactile cues were responsible for this increase in performance. To explore the role of tactile cues we ran another two studies. The first study focused on the tactile cues sensed by the pointing finger and the second on the tactile cues sensed by the palm.

6.4 STUDY 2: TACTILE CUES SENSED BY THE FINGER

In this study, we explored how much tactile cues sensed by the pointing finger contribute to interacting with a browsing-style imaginary interface. To test this, we compared three phone-based interaction surfaces: 1) a normal phone; 2) a phone with tactile cues added in the form of a tactile grid; 3) a phone with all cues removed by placing the interaction surface in a large featureless sheet of acrylic. To isolate the tactile cues we were primarily interested in blindfolded operation but also included sighted operation as an additional baseline.

6.4.1 Apparatus, task and procedure

The apparatus and task are identical to Study 1 except that the experimental conditions were changed. This study used a within-subjects 2×3 factorial design with these factors (shown in Figure 6.9):

SIGHTEDNESS: BLINDFOLDED VS. SIGHTED

INTERACTION SURFACE: PHONE VS. LARGE PHONE VS. TACTILE PHONE

We fabricated the phone prototypes in three layers: a 4 mm base of acrylic, a printed sheet of paper for phone screen and a 1.5 mm acrylic top layer, all glued together. The tactile grid on the surface of the phone used in the TACTILE PHONE condition (close-up shown in Figure 6.9d) was etched using a laser cutter, such that the etchings were approximately 0.4 mm deep. The interaction area of each phone prototype was identical (5×7.5 cm) and matched the interaction area on an iPhone 3G but for the LARGE PHONE (Figure 6.9b) the interaction area was centered on a 22.5×16.5 cm panel with rounded corners to prevent the participants from orienting using the device's bezel.

6.4.2 Hypotheses

By observing participants in pilot studies we noticed they regularly establish the location of the interaction surface by using the device's bezel when blindfolded. We therefore believe this is an important tactile cue

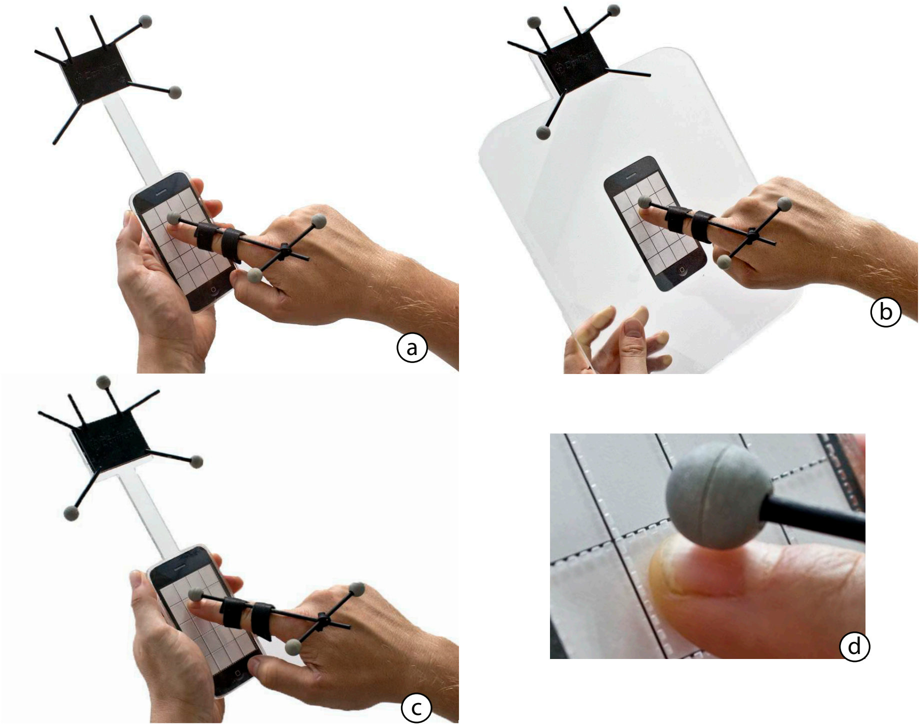


Figure 6.9: Study 2 conditions – (a) PHONE VS. (b) LARGE PHONE VS. (c) TACTILE PHONE; (d) close up of the tactile grid.

and we wished to confirm that depriving participants of it would result in worse performance.

HYPOTHESIS H1 When BLINDFOLDED, participants will be slower with the LARGE PHONE than with the PHONE.

Based on Study 1, where the palm, with its rich tactile cues, performed better than the smooth phone surface, we expected that adding tactile cues to the surface of the phone would also enable more efficient interaction.

HYPOTHESIS H2 When BLINDFOLDED, participants will be faster with the TACTILE PHONE than with the PHONE.

6.4.3 Participants

We recruited a new set of 12 participants from our institution (4 female, 10 right-handed). They were between the ages of 23 and 30 ($M = 25.2$, $SD = 2.55$) and all had normal or corrected to normal vision and hearing.

6.4.4 Results

From the 2943 data points collected during the experiment we removed 63 error trials (2.1%) and 72 outliers (2.4%), leaving 2808 trials for analysis. As in Study 1 outliers were defined as greater than three standard deviations above the mean per condition and repetition. The results were analyzed using repeated measures ANOVA. All post hoc comparisons used Bonferroni corrected confidence intervals. Each participant took approximately 45 minutes.

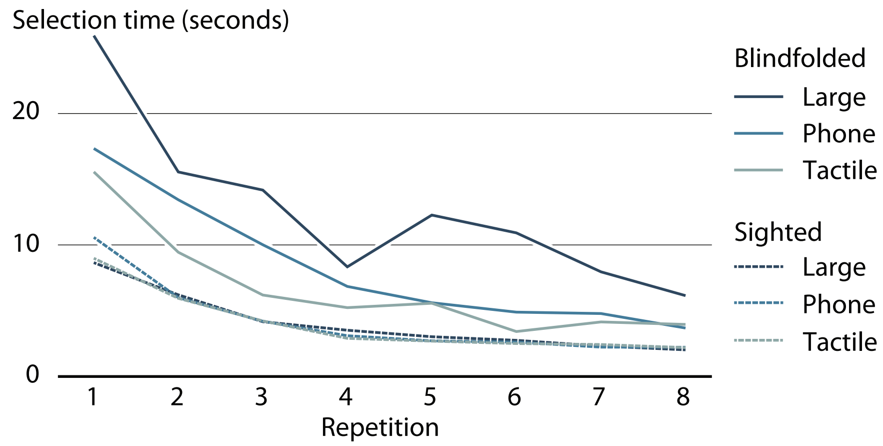


Figure 6.10: Study 2 performance of each condition over time.

The overall trend of the data, shown in Figure 6.10, matches the data from Study 1. We therefore took the same approach and divided the repetitions into two even groups of four repetitions each and performed separate analyses.

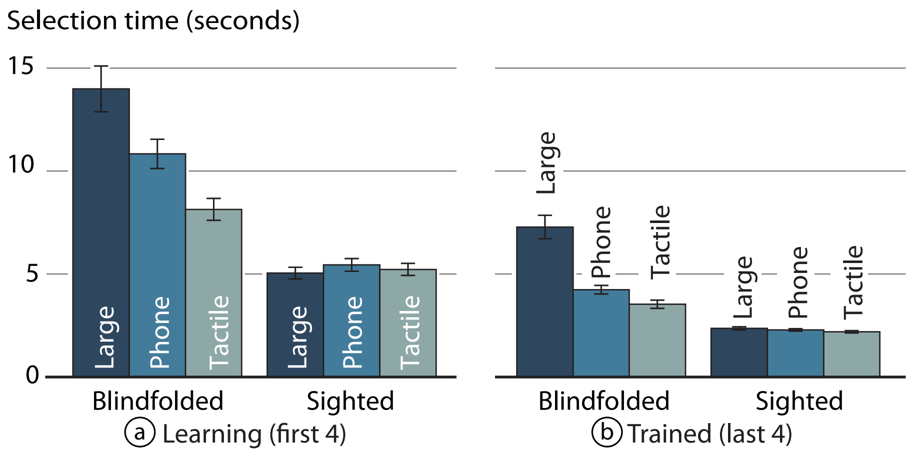


Figure 6.11: Study 2 aggregate selection times for the (a) first and (b) last half of the repetitions. Error bars are ± one standard error of the mean.

Learning phase (Figure 6.11a)

There was a significant main effect ($F_{2,22} = 11.51$, $p < 0.001$, $\eta^2 = 0.08$) of INTERACTION SURFACE and pairwise tests showed that participants were significantly faster with the TACTILE PHONE than both LARGE PHONE (30% faster, $p = 0.009$) and PHONE (18% faster, $p = 0.036$). Although the PHONE was faster than LARGE PHONE by 17%, the difference was not significant ($p = 0.061$).

When BLINDFOLDED participants were 110% slower than when they were SIGHTED ($F_{1,11} = 65.89$, $p < 0.001$, $\eta^2 = 0.48$) and there was a significant interaction between SIGHTEDNESS and INTERACTION SURFACE ($F_{2,22} = 12.21$, $p < 0.001$, $\eta^2 = 0.09$).

The differences between interaction surfaces were not significant when participants were SIGHTED ($p = 0.76$) but they were when participants were BLINDFOLDED ($F_{2,22} = 14.136$, $p < 0.001$, $\eta^2 = 0.36$), which explains the interaction effect. When BLINDFOLDED, participants using the TACTILE PHONE were significantly faster than both LARGE PHONE (42% faster, $p = 0.003$) and PHONE (25% faster, $p = 0.015$) but although the PHONE was faster by 23% than LARGE PHONE the difference was not significant $p = 0.060$).

Trained phase (Figure 6.11b)

As in the learning phase, there was significant main effect of INTERACTION SURFACE ($F_{2,22} = 18.18$, $p < 0.001$, $\eta^2 = 0.14$). Participants using the LARGE PHONE were significantly slower than when using the PHONE (48%, $p = 0.015$) and the TACTILE PHONE (41%, $p < 0.001$).

Overall in this phase, BLINDFOLDED participants were 120% slower than when SIGHTED ($F_{2,22} = 54.675$, $p < 0.001$, $\eta^2 = 0.36$). There was also a significant interaction between SIGHTEDNESS and INTERACTION SURFACE ($F_{2,22} = 15.99$, $p < 0.001$, $\eta^2 = 0.12$).

Breaking these results down further and looking at SIGHTED and BLINDFOLDED trials separately can help explain the interaction: when SIGHTED there was no significant difference between interaction surfaces ($p = 0.73$) but when BLINDFOLDED there was ($F_{2,22} = 17.836$, $p < 0.001$, $\eta^2 = 0.42$). Participants using the LARGE PHONE were significantly slower than those using the PHONE (72%, $p = 0.016$) and the TACTILE PHONE (106%, $p < 0.001$). The TACTILE PHONE was 17% faster than the PHONE but this difference was not significant ($p = 0.596$).

Overall in both phases and when BLINDFOLDED, the LARGE PHONE performed significantly worse than regular PHONE (and the TACTILE PHONE), which confirms our first hypothesis that depriving the participant of the bezel negatively affects performance. However, the TACTILE

PHONE only significantly improves interaction during the learning phase and not once the participants have learned the target locations. Therefore our second hypothesis regarding the benefits of added tactile cues is only partially confirmed.

6.4.5 Discussion

Our results suggest that, although a touchscreen phone appears featureless, some features do exist to guide interaction: for instance, the presence of a bezel provides a substantial benefit to the unsighted user. It allows users to find the extent of the interaction area and concentrate their exploration within that area. It also brings the participant's non-dominant hand near the interaction area, allowing it to be used as an additional tactile and proprioceptive cue.

Based on the results from Study 1 we expected that adding tactile cues to the surface of the phone would lead to a large improvement. However, this was not entirely the case: adding additional tactile cues only improved performance during the learning phase. Once the participants had learned where the targets were, they performed similarly with and without the extra tactile cues. This indicates that it is important to have some tactile features that can be sensed by the pointing finger but their advantage disappears after the user has learned the location of the interface elements.

However, in Study 1 we observed a large improvement during both phases of interaction when using the palm that was not observed with the extra tactile cues on the phone surface in this study. There is something beyond tactile features that set the palm apart from a phone's surface (even with added tactile features). In the next study, we look at the different types of tactile features that are available when using the palm to separate the role of active touch performed by the pointing finger from passive touch on the palm itself.

6.5 STUDY 3: TACTILE CUES SENSED BY THE PALM

To understand how important are the tactile cues sensed by the palm, this study removed the sensing of tactile cues from the palm. For comparison, we also added a condition that removed the sensing of cues by the pointing finger.

6.5.1 Apparatus, task and procedure

The apparatus and task are identical to Study 1 and 2 but, again, the interaction surface conditions were changed. This study used a within-subjects 2×3 factorial design with these factors (as shown Figure 6.12):

SIGHTEDNESS: SIGHTED VS. BLINDFOLDED

INTERACTION SURFACE: PALM VS. FAKE PALM VS. PALM WITH FINGER COVER.

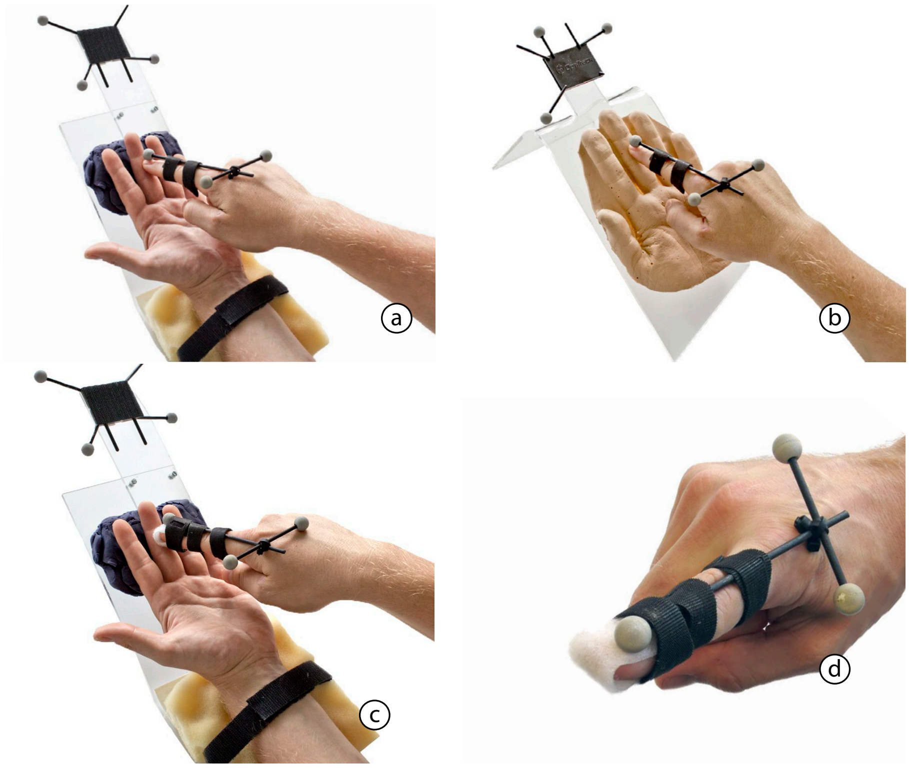


Figure 6.12: Study 3 conditions – (a) PALM VS. (b) FAKE PALM VS. (c) PALM WITH FINGER COVER; (d) close up of finger cover.

As in Study 1, for the PALM-based conditions we placed the participants' non-dominant hand in a fixture (shown in Figure 6.12a,c) that provided a consistent reference for calibration.

For the FAKE PALM condition, we built a realistic replica of my left hand (shown in Figure 6.12b) formed with liquid silicone. I used SORTA-Clear 40 liquid silicone mixed with skin colored silicone paint to create a replica that is slightly compliant and has a similar color and all of the fine ridges and features of a real hand. See Figure 6.13 for a comparison.

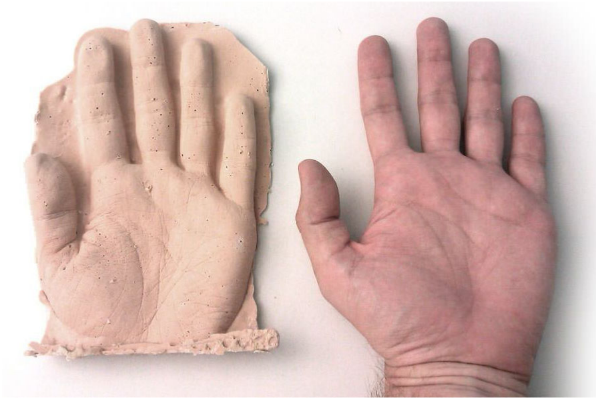


Figure 6.13: The silicone hand replica next to the author's hand.

For the PALM WITH FINGER COVER condition we covered the tip of the participant's pointing finger a piece of Velcro backing. The cover removed the fine cutaneous sensation from the participant's finger but the participants could still sense pressure and large features, like the palm outline.

6.5.2 Hypotheses

First, since the PALM condition allows the participants to use both palm and finger taction (i.e., intra-active touch (Bolanowski et al. 2004)) we expect it would outperform the other conditions when blindfolded:

HYPOTHESIS H1 When BLINDFOLDED, participants will be faster with the PALM than with the other interface conditions.

Secondly, we expect the FAKE PALM, which only involved active touch, to be comparatively worse to the PALM WITH FINGER COVER, which is dominated by passive touch. This is because the passive tactile discrimination on the palm is very good (Vallbo and Johansson 1978), which allows the participants to directly localize the sensation instead of integrating the position over time while scanning the interaction surface with the point finger. Therefore our second hypothesis is:

HYPOTHESIS H2 When BLINDFOLDED, participants will be slower with the FAKE PALM than with the PALM WITH FINGER COVER.

6.5.3 Participants

We recruited a new set of 12 participants from our institution (3 female, all right-handed) between the ages of 21 and 30 ($M = 24.3$, $SD = 2.67$). All had normal or corrected to normal vision and hearing.

6.5.4 Results

We collected 2941 data points and removed 61 error trials (2.0%) and 72 outliers (2.4%) using the same procedure was the other studies. This left 2808 trials for our analysis, which used the same procedure as Study 1 and 2. The performance over time is shown in Figure 6.14.

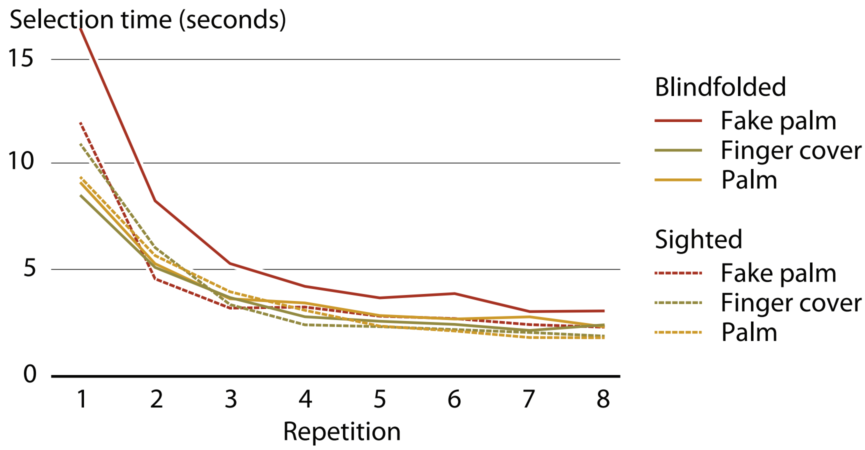


Figure 6.14: Study 3 performance for each condition over time.

Learning phase (Figure 6.15a)

There was no significant main effect of Interaction surface ($p = 0.083$) but there was for Sightedness: when blindfolded participants were 20% slower than when sighted ($F_{1,11} = 9.13$, $p = 0.012$, $\eta^2 = 0.07$). There was also an interaction effect between Interaction surface and Sightedness ($F_{2,22} = 10.54$, $p = 0.001$, $\eta^2 = 0.18$).

Only when blindfolded were the differences between interaction surfaces significant ($F_{2,22} = 8.145$, $p = 0.002$, $\eta^2 = 0.34$) with the fake palm being significantly slower than both the palm (38%, $p = 0.019$) and the palm with finger cover (33%, $p = 0.054$).

Trained phase (Figure 6.15b)

Unlike in the learning phase, this phase had a significant main effect of Interaction surface ($F_{2,22} = 12.08$, $p < 0.001$, $\eta^2 = 0.14$) with the fake palm being significantly slower than both the palm (25%, $p = 0.024$) and the palm with finger cover (24%, $p = 0.007$).

By breaking these numbers down further and only looking at blindfolded trials, there is still a significant main effect ($F_{2,22} = 12.173$, $p < 0.001$, $\eta^2 = 0.26$) and the differences are more pronounced with the fake palm being 30% slower than the palm ($p = 0.024$) and 33% slower than the palm with finger cover ($p = 0.003$).

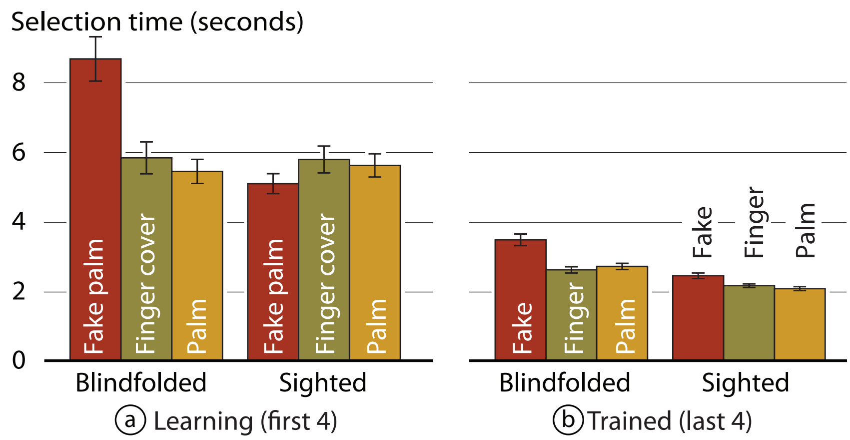


Figure 6.15: Study 3 aggregate selection times for the (a) first and (b) last half of the repetitions. Error bars are ± one standard error of the mean.

Overall, the FAKE PALM clearly performs the worst when BLINDFOLDED but surprisingly, for both interaction phases, the PALM and PALM WITH FINGER COVER conditions perform similarly, which partially confirms H1 (that PALM outperforms the others) and fully supports H2 (that FAKE PALM is slowest).

6.5.5 Discussion

The results indicate that it is the passive touch on the palm that contributes most to browsing an imaginary interface. The active touch feedback received by the tip of the pointing finger, in contrast, contributes comparatively little. Although the FAKE PALM condition contained equivalent tactile cues to the other conditions, participants performed substantially worse using it as an interaction surface.

We cannot say equivocally that the pointing finger contributes nothing to the interaction as in the PALM WITH FINGER COVER condition large-scale tactile features (such as the edges of the hand and fingers) could still be felt but it is apparent that the fine tactile cues on the surface of the palm contribute very little.

We believe the difference occurred because the high touch discriminability of the palm makes it inherently spatial—touch occurs at an easily resolvable location—whereas tactile cues sensed by the pointing finger are inherently ambiguous as all fingers provide similar tactile cues. Users are apparently able to resolve this by integrating tactile information over time to develop an understanding of where they are located on the palm. However, this integration process takes time and is prone to error, which would explain the longer interaction times in our studies.

The same reasoning can also explain the limited performance improvement of the TACTILE PHONE in Study 2. Since only the pointing finger could sense the added tactile cues, they contribute less than if the participant's palm could be used for sensing.

6.6 SUMMARY OF THE STUDIES

From these three studies we have gained an understanding of what enables palm-based imaginary interfaces. Specifically we discovered that:

1. Even though imaginary interfaces cannot display visual content, users' visual sense remains the main mechanism that allows users to control the interfaces because it allows users to watch their hands interact. In conditions where users are able to watch their hands interact, this overrides the other cues we studied, i.e., all tactile cues.
2. In the absence of visual cues, such as when driving or otherwise visually engaged, the tactile cues available when the pointing finger touches the palm can replace the lacking visual cues. As a result, palm-based imaginary interfaces remain usable even when operated eyes-free.
3. While we initially expected the pointing finger to sense the majority of tactile cues, we found the opposite to be the case, as the passive tactile sensing by the palm allows users to feel exactly where they are being touched. The most likely explanation is that the cues sensed by the pointing finger are ambiguous, while the cues sensed by the palm are unique and easy to locate spatially.

6.7 IMPLICATIONS FOR BLIND AND EYES-FREE USE

Our finding that tactile cues between pointing finger and supporting hand can, in part, replace the absent visual cues has an important implication: it suggests that using palm as an interaction surfaces (or even other parts of the body) provides not only a level of convenience beyond a dedicated interactive device but also has the potential to significantly improve the level of performance. We showed that when the interaction surface is located on the user's palm the additional passive tactile sensing increases performance (in our case by 33%) compared to an eyes-free interface on an ordinary surface (such as a mobile phone).

Although we only tested interaction on the palm, the sensing properties of human skin are available all over the body. While the palm is the most sensitive section of skin (Vallbo and Johansson 1978), there is an opportunity to exploit these capabilities on other parts of the body for eyes-free interaction. For example, a user's thigh is a readily available surface when seated in a train or bus, which could be used for text entry or general interaction with an audio-based interface.

The most direct use of this interaction could be for visually impaired users. While this project started by borrowing from the related work on interfaces for visually impaired users, we propose exporting our findings back to that community. More concretely, the Imaginary Interface hardware, e.g., sensing the hands with a chest-mounted camera, might allow visually impaired users to perform better than with the touchscreen-based devices they use today. While such a claim obviously requires a substantial amount of additional research, we want to conclude this chapter with a one-user pilot study we conducted to inspire this discussion.

One blind participant performing the task from Study 1

We recruited one blind person to perform the experiment task of Study 1 and to supply feedback.

Our participant was a 33 year-old male, right-handed and a musician by trade. He had been blind since age two and has zero sensitivity to light. In his daily life, he uses screen-reading software on his PC and on his non-touchscreen Nokia mobile phone. He was familiar with the VoiceOver for iPhone interaction style but had not used it regularly.

The participant performed the task from Study 1. He performed four blocks (two for each interaction surface condition) of 40 trials each. We used ABBA counterbalancing to balance learning effects. His results are shown in Figure 6.16.

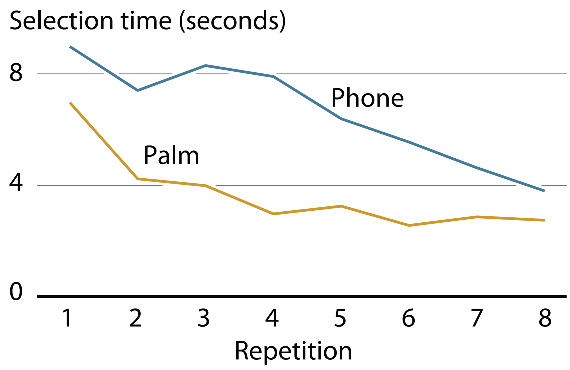


Figure 6.16: Blind participant's selection times with the PALM and PHONE interfaces.

Overall, his performance matched the results from the blindfolded participants in that he was faster in both phases with the palm interface. In the learning phase he was 44% faster with the palm (4.54 sec.) than phone (8.15 sec.) and also 44% faster in the trained phase (2.85 sec. vs. 5.09 sec.).

Following the study, we conducted an informal interview. He was overall very positive about the palm interface and preferred it to the phone, saying that he preferred "the material of [his] palm." Assuming the sensing technology was reliable, he said that he could imagine himself using such an interface. He also commented that using the palm might actually have less social stigma in public because it wouldn't appear out of the ordinary, especially compared to specialized equipment like Braille readers.

Clearly we must be careful generalizing from the outcome of one participant but the results here are promising and will hopefully inspire future work in the area of imaginary interfaces for visually impaired users.

CONCLUSIONS

In this chapter, we sought to gain a deeper understanding of what processes enable palm-based imaginary interfaces and explored which inherent properties of these interfaces are responsible for user performance. We conducted our exploration using the browsing interface, which uses audio feedback to enable operation of an unfamiliar imaginary interface. We learned that there are three central feedback channels that enable the interaction: first, visual cues are the strongest, followed by tactile cues sensed by the palm and finally, tactile cues sensed by the pointing finger.

Although these results tell us mostly about the properties of Imaginary Interfaces, they also have two related implications for other non-visual interactive systems. Since the passive taction of the palm provides a large amount of the spatial touch discrimination we argue that palm-based interaction (and perhaps other locations of the body) has the potential to significantly improve eyes-free performance compared to interaction on a dedicated device (such as a mobile phone). The most applicable application of this idea is in the designing of interfaces for visually impaired users; using a palm-based imaginary interface has the potential to significantly improve performance over the current state of the art interaction style (VoiceOver for iPhone).

7 CONCLUSIONS AND FUTURE WORK

As I stated in the introduction, the goal of this research was to create an ultra-mobile interface that retains the spatial interaction style from traditional mobile devices.

With Imaginary Interfaces, I have shown that it is indeed possible to create such an interface, albeit with some compromises. I have shown that by exploiting users' visuospatial memory, they are able to extend, annotate or edit a drawing after it was initially created, standing still or while mobile. I have shown that by using the palm as interaction surface, users are able to transfer the knowledge they have of another interface to learn and operate an imaginary interface. Finally, I have shown that by supplying audio feedback, users are able to explore and learn previously unknown interfaces.

Since Imaginary Interfaces moves the interaction surface from a dedicated physical touchscreen to the user's environment, the interfaces are not limited by the size of the touchscreen, as current mobile devices are. They can be made as small as the sensing element, which is currently in a stage of rapid miniaturization. However, unlike other ultra-mobile interfaces, Imaginary Interfaces retain the spatial interaction style from traditional devices, which allows users to transfer their skills and knowledge to this new interface.

7.1 LIMITATIONS

Although, I stand strongly behind the conclusions derived from this work, there are several limitations and areas where support could have been stronger.

The prototypes presented in Chapter 4 exist only to showcase the interaction styles and to suggest how a deployable interface might be built. However, the engineering behind these prototypes has not been fine tuned to the point where it is possible to test these prototypes outside of the lab.

The form factor comparison study in Section 4.3 used "perfect" tracking (i.e., post hoc analysis of high resolution photographs) which does a good job of measuring raw human performance, but is not a measure of the accuracy users would expect with a realistic (i.e., imperfect) tracking mechanism. A more representative measure would need to take into account the noisy sensing capabilities of a deployed sensing platform and the effect of distraction and changes in the environment during interaction.

In most of the studies presented in this dissertation, we concentrated on target selection only. However, interaction with touchscreen devices also involves dragging, swiping and multi-finger gestures (such as pinch-to-zoom), which we have not yet fully explored.

Also, it is possible that the participants we used in the studies were not representative of the general population. They tended to be young, technology-minded people, often male students in a computer science department. It is possible that some conclusions may not hold for other segments of the population.

In Chapter 5, I presented the Imaginary Phone as a shortcut method for common phone functions. However, we only looked at the participant's ability to remember the icons on their home screen. Real world use of an interface like this would required knowledge of much more complex sequence of interaction events, such as the multi-step sequence require to place a phone call to a previously saved contact. While our studies indicate that this might be possible, I do not provide any explicit support for multi-step interactions.

The cafeteria study in Section 5.3 has limited predictive power because of the small number of participants (only six per condition) relative to the variability of participants' responses. Also, by only sampling students that owned iPhones, we further increase the bias towards toward young, college-aged users who were especially technology savvy. Other populations might exhibit different amounts of inadvertent spatial learning.

7.2 FUTURE WORK

With the investigations and designs presented in this dissertation I hope to have laid the groundwork for mobile interfaces that are both spatial and non-visual. From here I can imagine several directions for continued research.

First, since knowing where UI elements exist in an imaginary interface is a fundamental issue, there is clear motivation for improving upon the techniques proposed in this dissertation. For instance, perhaps an imaginary interface could be learned by watching someone use the interface. Just like users obtain sufficient visuospatial memory by watching their own interaction, a user should be able to do the same by closely watching someone else. The area of collaborative interactions with Imaginary Interfaces has many open questions and possibilities for future work.

Currently all imaginary interfaces presented in this dissertation are based on a flat plane of interaction. There is potential to increase the expressiveness of Imaginary Interfaces by extending them to 3D or onto non-planar surfaces, although care must be taken to ensure that the added expressiveness of 3D interaction is not overshadowed by the well-known issues of learnability with 3D interfaces (Cockburn and McKenzie 2002).

Chapter 6 showed that the tactile sensing available on the user's palm can significantly improve interaction performance. This idea could be expanded to include other parts of the body. Wagner et al. (2013) provides a design space and initial investigation into the use of the body to supplement more conventional interaction styles, which also motivates more work in this area.

Finally, the ideas presented in this dissertation could also be explored beyond interaction with mobile devices. Transfer learning especially, could be applied to a broader range of devices, such as remote controls and instrument panels. It would be interesting to investigate if the transfer learning principle could be applied to these devices (which have a strong tactile component) rather than the visual interfaces as I have shown here. For instance, Imaginary Devices (Steins et al. 2013) uses the principle of transfer learning to help users learn the gestures required to operate 'imaginary' versions of traditional input devices (e.g., joysticks and steering wheels) and Imaginary Reality Gaming (Baudisch et al. 2013) allows players to transfer their knowledge of a physical sport (e.g., football or basketball) to play a new type of game where players must imagine the location of the ball and operate on a shared reality amongst the other players.

7.3 FINAL WORDS

With this body of work, I hope to do more than to present the design of a new style of interaction. I hope to convince my readers that vision need not be the primary interface between the user and a mobile device. By exploiting users' spatial memory and by supplying liberal amounts of audio feedback, combined with the passive sensing on the users' palm, it is possible to create a very functional mobile device that does not rely of visual feedback. In other words, I found that much more is possible with a non-visual mobile interface than is generally thought possible (at I least I did not).

In particular, I would like to inspire future researchers and product designers to accept the difficult task of designing interfaces, interactions and devices that are not so greedy with our visual channel. All too often today, people are visually (and hence cognitively) engrossed in their mobile devices. These mobile devices have come to *replace* our experiences with the real world and unfortunately fail in their promise to augment and improve our lives. I believe it is possible to create a more humane interface that exploits other sensory channels and leaves the user's attention unencumbered, so they may more fully experience the real world.

I believe in the words of Stuart Card:

> *"We should be careful to make the world we actually want to live in."*

And I ask you, dear reader, to considering doing the same. To take the lessons learned from this dissertation and produce the next generation of mobile devices that avoids the trap of creating sealed-off immersive experiences that take us away from what is happening right in front of us and follow through with technology's promise of augmenting our lives for the better.

Appendix

A SURVEY OF MOST COMMON APPS

To obtain an ecologically valid basis for my phone-like interfaces, I conducted a survey of iPhone-owning students in my institution. Through a broadcast email, I recruited 47 students that returned screenshots of the home screen on their iPhones.

From these screenshots I compiled a list of the most common apps on the home screen of these devices. This list was used to create an ecologically valid list of application names for the studies in this dissertation.

Place	Name	Count	Place	Name	Count
1	Mail	46	20	Skype	11
2	Safari	44	21	Stocks	10
3	Calendar	43	22	iTunes	9
4	Phone	42	23	Twitter	8
5	Settings	40	24	Reeder	6
6	Messages	39	25	WeatherPro	5
7	Maps	38	26	Voice Memos	5
8	Camera	36	27	dict.cc	4
9	iPod	35	28	Things	4
10	Photos	35	29	Sleep Cycle	4
11	Clock	34	30	Game Center	4
12	Notes	26	31	Dropbox	4
13	Contacts	25	32	WhatsApp	3
14	Facebook	22	33	Videos	3
15	Weather	21	34	Shazam	3
16	Calculator	21	35	NAVIGON	3
17	App Store	20	36	Instapaper	3
18	YouTube	19	37	BeejiveIM	3
19	FahrInfo	18	38	Articles	3

B SELECTED EXPERIMENT MATERIALS

The following pages contain questionnaires and other experimental material from the studies reported in this dissertation.

B.1 PALM VS. MIDAIR STUDY

Questionnaire from study in Section 4.3.

Page **1** of **2**

Learning Imaginary Interfaces Study – Questionnaire and Consent Form

Date: ______________________

Name: ______________________________

Age: _________

Gender (circle one): Male Female

Preferred Hand (circle one): Left Right

I agree to participate in this study.

Signature: _________________________

Instructions:

This experiment has two parts. In the first your task is to completely learn the location of three targets on the photograph of your hand. In the second you will be asked to select those targets on your own hand with the highest accuracy possible.

Now take a close look at these three targets. Spend 30 seconds mentally recording exactly where they are. Look for nearby landmarks, lines, creases, finger tips, etc. It is important that you learn the positions exactly as possible because in the second part of the study we want to know how accurately you can select these positions on your actual hand.

Now follow the directions on-screen and select the targets on the iPod as they are announced.

When complete < flip over > ...

Imagine yourself in 5 years. You just purchased the newest iPhone version and it comes with two modes of operation: 1) the traditional touchscreen mode and 2) a 'hands-free' mode where you can leave the phone in your pocket and still perform simple operations like answer a call or pause the music player. When in 'hands-free' mode you use either the palm of your hand as an interaction surface or the empty space framed by an L gesture.

Based on what you have just done in the experiment which of these two interaction surfaces would you prefer and why?

Other Comments:

B.2 TRANSFER LEARNING STUDY

Questionnaire, experimenter worksheet and explanatory materials from study in Section 5.3.

iPhone Recall Study

Informed Consent

During this study we will ask questions about your iPhone use and we will take a photo of the icons on your phone. You have the right to withdraw from the study at anytime.

I agree to participate in this study: Name: ______________________________

Signature: __________________________ Date: ____________________

Questionnaire

Age: _________ Gender (circle one): Male Female

Which hand do you write with (circle one): Left Right

How long have you been using an iPhone? (circle one) less than 1 month

from 1 month to one year

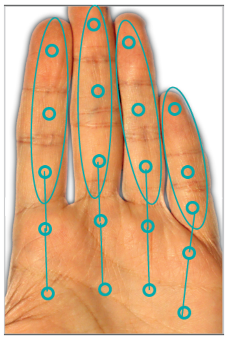
greater than 1 year

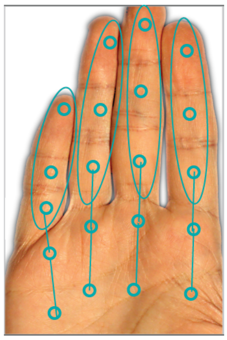
Comments:

INTERNAL USE	Participant ID:	Group:	(h=iPhone, t=hand)

iPhone Recall Study

draw arrow FROM correct TO incorrect position





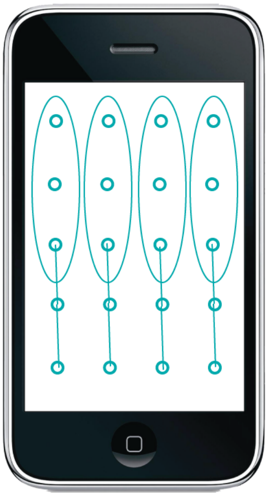
How often used:

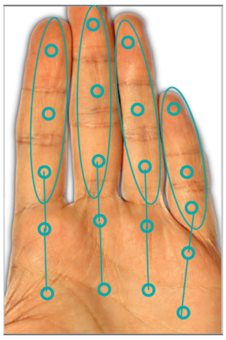
- D – daily
- W – at least once a week
- R – rarely

Participant ID:

iPhone Recall Study

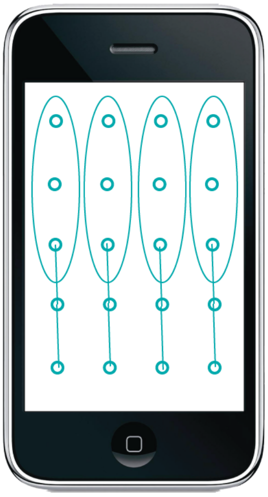
Right Handed Users

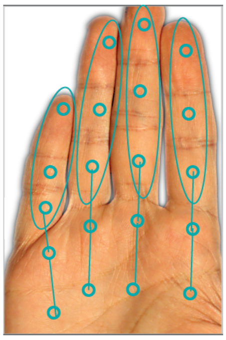




iPhone Recall Study

Left Handed Users





B.3 SELECTION STUDY

Questionnaire from study in Section 5.4.

Hand Interaction Study

Informed Consent

During this study you will operate a non-visual phone prototype. We will record your interactions. You have the right to withdraw from the study at anytime.

I agree to participate in this study: Name: ______________________________

Signature: ____________________________ Date: ____________________

Questionnaire

Age: ________ Gender (circle one): Male Female

Which hand do you write with (circle one): Left Right

< Flip over after the study and answer a couple more questions. >

INTERNAL USE	Participant ID:	Condition:

Hand Interaction Study

When blind which interface did you prefer to use (circle one):

Palm iPod with no handle iPod with the handle

Why:

When not blind which interface did you prefer to use (circle one):

Palm iPod with no handle iPod with the handle

Why:

Other Comments:

B.4 ROLE OF VISUAL CUES STUDY

Questionnaire from study in Section 6.3.

Blind Hand Interaction Study

Informed Consent

During this study you will operate a non-visual phone prototype. For part of the study you will be partially blindfolded.

We may record (through instrumentation, photography, audio and/or video) your interactions during the study. We reserve the right to publicly present and publish non-identifying details, images, video and quotes captured during your participation in this study.

You have the right to withdraw from the study at anytime without penalty.

I agree to participate in this study: Name: ______________________________

Signature: __________________________ Date: __________________

Questionnaire

Age: ________ Gender (circle one): Male Female

Which hand do you write with (circle one): Left Right

< After the study: flip over and answer a few more questions. >

INTERNAL USE	Participant ID:	Condition:

Blind Hand Interaction Study

When blindfolded:

Which was easier to use? (circle one)	Hand	Phone	Neither
Which was faster to use? (circle one):	Hand	Phone	Neither
Which did you prefer to use? (circle one)	Hand	Phone	Neither

When *not* blindfolded:

Which was easier to use? (circle one)	Hand	Phone	Neither
Which was faster to use? (circle one):	Hand	Phone	Neither
Which did you prefer to use? (circle one)	Hand	Phone	Neither

Other Comments:

BIBLIOGRAPHY

Ahmad, F. and Musilek, P. (2006). A keystroke and pointer control input interface for wearable computers. In *Proceedings of PerCom*, pages 2–11 (cited on page 16).

Alibali, M. W. (2005). Gesture in spatial cognition: expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation* 5.4, pages 307–331.

Amma, C., Georgi, M. and Schultz, T. (2013). Airwriting: a wearable handwriting recognition system. *Personal and Ubiquitous Computing*, pages 1–13 (cited on page 14).

Ängeslevä, J., Oakley, I., Hughes, S. and O'Modhrain, S. (2003). Body Mnemonics portable device interaction design concept. In *Adjunct Proceedings of UIST (Poster)*, pages 2–5 (cited on page 12).

Apple Inc. (2012). *iOS Human Interface Guidelines*. Last updated: 2012-12-17 (cited on page 80).

Ashbrook, D., Baudisch, P. and White, S. (2011). Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of CHI*, pages 2043–2046 (cited on page 14).

Ashbrook, D., Clawson, J., Lyons, K., Starner, T. and Patel, N. (2008). Quickdraw: the impact of mobility and on-body placement on device access time. In *Proceedings of CHI*, pages 219–222 (cited on pages 14, 65).

Ashbrook, D., Lyons, K. and Starner, T. (2008). An investigation into round touchscreen wristwatch interaction. In *Proceedings of MobileHCI*, pages 311–314 (cited on page 11).

Ashbrook, D. (2010). Enabling Mobile Microinteractions. PhD thesis. Georgia Institute of Technology (cited on pages 14, 65).

Baddeley, A. D. and Hitch, G. (1974). Working memory. In *The Psychology of Learning and Motivation: Advances in Research and Theory*. Edited by G. Bower. Volume 8, pages 47–89 (cited on page 20).

Bailly, G., Müller, J., Rohs, M., Wigdor, D. and Kratz, S. (2012). Shoesense: a new perspective on gestural interaction and wearable applications. In *Proceedings of CHI*, pages 1239–1248 (cited on page 16).

Balakrishnan, R. and Hinckley, K. (1999). The role of kinesthetic reference frames in two-handed input performance. In *Proceedings of UIST*, pages 171–178 (cited on page 8).

Baudel, T. and Beaudouin-Lafon, M. (1993). Charade: remote control of objects using free-hand gestures. *Communications of the ACM* 36.7, pages 28–35 (cited on page 18).

Baudisch, P. and Chu, G. (2009). Back-of-device interaction allows creating very small touch devices. In *Proceedings of CHI*, pages 1923–1932 (cited on pages 10, 11).

Baudisch, P., Pohl, H., Reinicke, S., Wittmers, E., Lühne, P., Knaust, M., Köhler, S., Schmidt, P. and Holz, C. (2013). Imaginary reality gaming: ball games without a ball. In *Proceedings of UIST*, pages 405–410 (cited on pages 10, 107).

Beaudouin-Lafon, M. (2000). Instrumental interaction: an interaction model for designing post-WIMP user interfaces. In *Proceedings of CHI*, pages 446–453 (cited on page 17).

Berkinblit, M. B., Fookson, O. I., Smetanin, B., Adamovich, S. V. and Poizner, H. (1995). The interaction of visual and proprioceptive inputs in pointing to actual and remembered targets. *Experimental Brain Research* 107.2, pages 326–330 (cited on page 20).

Billinghurst, M., Bowskill, J., Dyer, N. and Morphett, J. (1998). Spatial information displays on a wearable computer. *IEEE Computer Graphics and Applications* 18.6, pages 24–30 (cited on page 8).

Blaskó, G., Coriand, F. and Feiner, S. (2005). Exploring interaction with a simulated wrist-worn projection display. In *Proceedings of ISWC*, pages 2–9 (cited on page 12).

Blaskó, G. and Feiner, S. (2004). An interaction system for watch computers using tactile guidance and bidirectional segmented strokes. In *Proceedings of ISWC*, pages 120–123 (cited on page 11).

Bolanowski, S. J., Verrillo, R. T. and McGlone, F. (2004). Passive, active and intra-active (self) touch. *Behavioural Brain Research* 148, pages 41–45 (cited on pages 21, 99).

Bolt, R. A. (1980). "Put-that-there": voice and gesture at the graphics interface. In *Proceedings of SIGGRAPH*, pages 262–270 (cited on page 18).

Brewster, S., Lumsden, J., Bell, M., Hall, M. and Tasker, S. (2003). Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of CHI*, pages 473–480 (cited on page 19).

Butler, A., Izadi, S. and Hodges, S. (2008). SideSight: multi-"touch" interaction around small devices. In *Proceedings of UIST*, pages 201–204 (cited on page 9).

Cao, X. and Balakrishnan, R. (2006). Interacting with dynamically defined information spaces using a handheld projector and a pen. In *Proceedings of UIST*, pages 225–234 (cited on page 9).

Cao, X., Forlines, C. and Balakrishnan, R. (2007). Multi-user interaction using handheld projectors. In *Proceedings of UIST*, pages 43–52 (cited on page 9).

Chen, X., Marquardt, N., Tang, A., Boring, S. and Greenberg, S. (2012). Extending a mobile device's interaction space through body-centric interaction. In *Proceedings of MobileHCI*, pages 151–160 (cited on page 14).

Cockburn, A., Quinn, P., Gutwin, C., Ramos, G. and Looser, J. (2011). Air pointing: design and evaluation of spatial target acquisition with and without visual feedback. *International Journal of Human-Computer Studies* 69.6, pages 401–414 (cited on page 8).

Cockburn, A. and McKenzie, B. (2002). Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *Proceedings of CHI*, pages 203–210 (cited on pages 7, 107).

Crossan, A., McGill, M., Brewster, S. and Murray-Smith, R. (2009). Head tilting for interaction in mobile contexts. In *Proceedings of MobileHCI*, article no. 6 (cited on page 15).

Czerwinski, M., Dantzich, M. van, Robertson, G. G. and Hoffman, H. (1999). The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3D. In *Proceedings of INTERACT*, pages 163–170 (cited on page 7).

Darling, W. G. and Miller, G. F. (1993). Transformations between visual and kinesthetic coordinate systems in reaches to remembered object locations and orientations. *Experimental Brain Research* 93.3, pages 534–547 (cited on page 20).

Deen, J., Lee, S. C., Li, B. H. and Starner, T. (2010). Improving the form factor of a wrist-based mobile gesture interface. In *Extended Abstracts of CHI (Poster)*, pages 3679–3684 (cited on page 11).

Dezfuli, N., Khalilbeigi, M., Huber, J., Müller, F. and Mühlhäuser, M. (2012). PalmRC: imaginary palm-based remote control for eyes-free television interaction. In *Proceedings of EuroiTV*, pages 27–34 (cited on page 13).

Driver, J. and Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences* 2.7, pages 254–262 (cited on page 21).

Emmorey, K., Tversky, B. and Taylor, H. A. (2000). Using space to describe space: perspective in speech, sign, and gesture. *Spatial Cognition and Computation* 2.3, pages 157–180 (cited on page 18).

Ens, B., Ahlström, D., Cockburn, A. and Irani, P. (2011). Characterizing user performance with assisted direct off-screen pointing. In *Proceedings of MobileHCI*, pages 485–494 (cited on page 9).

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences* 8.4, pages 162–169 (cited on page 21).

Fallah, N., Apostolopoulos, I., Bekris, K. and Folmer, E. (2012). The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks. In *Proceedings of CHI*, pages 425–432 (cited on page 10).

Fitzmaurice, G. W. (1993). Situated information spaces and spatially aware palmtop computers. *Communications of the ACM* 36.7, pages 39–49 (cited on page 9).

Folmer, E. and Morelli, T. (2012). Spatial gestures using a tactile-proprioceptive display. In *Proceedings of TEI*, pages 139–142 (cited on pages 10, 14).

Fuentes, C. T. and Bastian, A. J. (2010). Where is your arm? Variations in proprioception across space and tasks. *Journal of Neurophysiology* 103, pages 164–171 (cited on pages 21, 89).

Fukumoto, M. and Tonomura, Y. (1999). Whisper: a wristwatch style wearable handset. In *Proceedings of CHI*, pages 112–119 (cited on pages 11, 12).

Fukumoto, M. and Tonomura, Y. (1997). "Body coupled FingeRing": wireless wearable keyboard. In *Proceedings of CHI*, pages 147–154 (cited on page 11).

Gibson, J. J. (1962). Observations on active touch. *Psychological Review* 69.6, pages 477–491 (cited on page 21).

El-Glaly, Y. N., Quek, F., Smith-Jackson, T. and Dhillon, G. (2013). Touchscreens are not tangible: fusing tangible interaction with touch glass in readers for the blind. In *Proceedings of TEI*, pages 245–253.

Goldstein, M. and Chincholle, D. (1999). Finger-joint gesture wearable keypad. In *Proceedings of MobileHCI*, pages 9–18 (cited on page 13).

Gollner, U., Bieling, T. and Joost, G. (2012). Mobile lorm glove: introducing a communication device for deaf-blind people. In *Proceedings of TEI*, pages 127–130 (cited on pages 13, 15).

Greiner, T. M. (1991). *Hand Anthropometry of U.S. Army Personnel*. U.S. Army Report ADA244533 (cited on pages 40, 54, 76).

Guerreiro, T., Gamboa, R. and Jorge, J. (2008). Mnemonical body shortcuts: improving mobile interaction. In *Proceedings of ECCE*, 11:1–11:8 (cited on page 12).

Guiard, Y. (1987). Asymmetric division of labor in human skilled bimanual action: the kinematic chain as a model. *Journal of Motor Behaviour* 19.4, pages 486–517 (cited on page 8).

Gustafson, S. (2012). Imaginary Interfaces: touchscreen-like interaction without the screen. In *Extended Abstracts of CHI (Doctoral Consortium)*, pages 927–930 (cited on page vii).

Gustafson, S., Bierwirth, D. and Baudisch, P. (2010). Imaginary Interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of UIST*, pages 3–12 (cited on page vii).

Gustafson, S., Holz, C. and Baudisch, P. (2011). Imaginary Phone: learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of UIST*, pages 283–292 (cited on page vii).

Gustafson, S., Rabe, B. and Baudisch, P. (2013). Understanding palm-based imaginary interfaces: the role of visual and tactile cues when browsing. In *Proceedings of CHI*, pages 889–898 (cited on page vii).

Guth, D. (1990). Space saving statistics: an introduction to constant error, variable error, and absolute error. *Peabody Journal of Education* 67.2, pages 110–120 (cited on page 54).

Hännikäinen, J., Mikkonen, J. and Vanhala, J. (2005). Button component encasing for wearable technology applications. In *Proceedings of ISWC*, pages 205–206 (cited on page 58).

Harrison, C. and Hudson, S. E. (2009). Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proceedings of UIST*, pages 121–124 (cited on pages 9, 11).

Harrison, C., Ramamurthy, S. and Hudson, S. E. (2012). On-body interaction: armed and dangerous. In *Proceedings of TEI*, pages 69–76 (cited on page 12).

Harrison, C., Benko, H. and Wilson, A. D. (2011). OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of UIST*, pages 441–450 (cited on pages 12, 16, 17).

Harrison, C. and Hudson, S. E. (2010). Minput: enabling interaction on small mobile devices with high-precision, low-cost, multipoint optical tracking. In *Proceedings of CHI*, pages 1661–1664 (cited on page 9).

Harrison, C., Tan, D. and Morris, D. (2010). Skinput: appropriating the body as an input surface. In *Proceedings of CHI*, pages 453–462 (cited on pages 12, 15).

Hasan, K., Ahlström, D. and Irani, P. (2013). AD-Binning: leveraging around device space for storing, browsing and retrieving mobile device content. In *Proceedings of CHI*, pages 899–908 (cited on page 9).

Hinckley, K., Pausch, R. and Proffitt, D. (1997). Attention and visual feedback: the bimanual frame of reference. In *Proceedings of Symposium on Interactive 3D Graphics*, pages 121–126 (cited on pages 8, 33).

Holz, C. and Baudisch, P. (2010). The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *Proceedings of CHI*, pages 581–590 (cited on pages 10, 75, 80, 82).

Holz, C., Grossman, T., Fitzmaurice, G. and Agur, A. (2012). Implanted user interfaces. In *Proceedings of CHI*, pages 503–512 (cited on page 12).

Holz, C. and Wilson, A. (2011). Data Miming: inferring spatial object descriptions from human gesture. In *Proceedings of CHI*, pages 811–820 (cited on page 10).

Hornof, A., Halverson, T., Isaacson, A. and Brown, E. (2008). Transforming object locations on a 2D visual display into cued locations in 3D auditory space. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, pages 1170–1174 (cited on page 20).

Hrabia, C.-E., Wolf, K. and Wilhelm, M. (2013). Whole hand modeling using 8 wearable sensors: biomechanics for hand pose prediction. In *Proceedings of Augmented Human*, pages 21–28 (cited on page 16).

Hudson, S. E., Harrison, C., Harrison, B. L. and LaMarca, A. (2010). Whack Gestures: inexact and inattentive interaction with mobile devices. In *Proceedings of TEI*, pages 109–112 (cited on page 14).

Jackson, B. and Keefe, D. F. (2011). Sketching over props: understanding and interpreting 3d sketch input relative to rapid prototype props. In *Proceedings of IUI 2011 Sketch Recognition Workshop*, 6 pages. URL: `http://ivlab.cs.umn.edu/papers/Jackson-2011-SketchRecognition.pdf` (visited on 04/16/2013) (cited on page 18).

Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T. and Zigelbaum, J. (2008). Reality-based interaction: a framework for post-WIMP interfaces. In *Proceedings of CHI*, pages 201–210 (cited on page 17).

Jones, B., Sodhi, R., Forsyth, D., Bailey, B. and Maciocci, G. (2012). Around device interaction for multiscale navigation. In *Proceedings of MobileHCI*, pages 83–92 (cited on page 9).

Kajastila, R. and Lokki, T. (2013). Eyes-free interaction with free-hand gestures and auditory menus. *International Journal of Human-Computer Studies* 71.5, pages 627–640 (cited on page 10).

Kane, S. K., Jayant, C., Wobbrock, J. O. and Ladner, R. E. (2009). Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of ASSETS*, pages 115–122 (cited on page 19).

Kane, S. K., Bigham, J. P. and Wobbrock, J. O. (2008). Slide Rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of ASSETS*, pages 73–80 (cited on pages 19, 84).

Karam, M. and Schraefel, M. C. (2005). *A taxonomy of gestures in human computer interactions*. Technical Report ECSTR-IAM05-009. Electronics and Computer Science, University of Southampton (cited on page 18).

Kasahara, S., Heun, V., Lee, A. S. and Ishii, H. (2012). Second surface: multi-user spatial collaboration system based on augmented reality. In *Proceedings of SIGGRAPH Asia Emerging Technologies*, article no. 20 (cited on page 9).

Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I. and Olivier, P. (2012). Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of UIST*, pages 167–176 (cited on page 16).

Kim, J., He, J., Lyons, K. and Starner, T. (2007). The Gesture Watch: a wireless contact-free gesture based wrist interface. In *Proceedings of ISWC*, pages 15–22 (cited on page 11).

Kim, S., Takahashi, S. and Tanaka, J. (2010). New interface using palm and fingertip without marker for ubiquitous environment. In *Proceedings of ICIS*, pages 819–824 (cited on page 12).

Kohli, L. and Whitton, M. (2005). The Haptic Hand: providing user interface feedback with the non-dominant hand in virtual environments. In *Proceedings of GI*, pages 1–8 (cited on page 13).

Kolb, A., Barth, E., Koch, R. and Larsen, R. (2010). Time-of-flight cameras in computer graphics. *Computer Graphics Forum* 29.1, pages 141–159 (cited on page 48).

Kratz, S. and Rohs, M. (2009). HoverFlow: expanding the design space of around-device interaction. In *Proceedings of MobileHCI*, 4:1–4:8 (cited on page 9).

Kristoffersen, S. and Ljungberg, F. (1999). "Making place" to make IT work: empirical explorations of HCI for mobile CSCW. In *Proceedings of GROUP*, pages 276–285 (cited on page 12).

Krueger, M. W., Gionfriddo, T. and Hinrichsen, K. (1985). VIDEOPLACE—an artificial reality. In *Proceedings of CHI*, pages 35–40 (cited on page 18).

Kuester, F., Chen, M., Phair, M. E. and Mehring, C. (2005). Towards keyboard independent touch typing in VR. In *Proceedings of VRST*, pages 86–95 (cited on pages 13, 15, 48).

Làdavas, E., Farnè, A., Zeloni, G. and Pellegrino, G. di (2000). Seeing or not seeing where your hands are. *Experimental Brain Research* 131, pages 458–467 (cited on pages 21, 89).

Landua, S. and Wells, L. (2003). Merging tactile sensory input and audio data by means of the Talking Tactile Tablet. In *Proceedings of Eurohaptics*, pages 414–418 (cited on page 19).

Lee, S. C. and Starner, T. (2009). Mobile gesture interaction using wearable tactile displays. In *Extended Abstracts of CHI (Poster)*, pages 3437–3442 (cited on page 11).

Li, F. C. Y., Dearman, D. and Truong, K. N. (2010). Leveraging proprioception to make mobile phones more accessible to users with visual impairments. In *Proceedings of ASSETS*, pages 187–194 (cited on page 10).

Li, F. C. Y., Dearman, D. and Truong, K. N. (2009). Virtual Shelves: interactions with orientation aware devices. In *Proceedings of UIST*, pages 125–128 (cited on page 10).

Li, K. A., Baudisch, P. and Hinckley, K. (2008). Blindsight: eyes-free access to mobile phones. In *Proceedings of CHI*, pages 1389–1398 (cited on page 19).

Lin, S.-Y., Su, Z.-H., Cheng, K.-Y., Liang, R.-H., Kuo, T.-H. and Chen, B.-Y. (2011). PUB - Point Upon Body: exploring eyes-free interactions and methods on an arm. In *Proceedings of UIST*, pages 481–488 (cited on pages 12, 15).

Loclair, C., Gustafson, S. and Baudisch, P. (2010). PinchWatch: a wearable device for one-handed microinteractions. In *Proceedings of Workshop on Ensembles of On-Body Devices*, 4 pages (cited on page 14).

Lucero, A., Keränen, J. and Jokela, T. (2010). Social and spatial interactions: shared co-located mobile phone use. In *Extended Abstracts of CHI (Poster)*, pages 3223–3228 (cited on page 9).

Mackenzie, S. and Zhang, S. (1997). The immediate usability of Graffiti. In *Proceedings of GI*, pages 129–137 (cited on page 28).

Mayol-Cuevas, W. W., Tordoff, B. J. and Murray, D. W. (2009). On the choice and placement of wearable vision sensors. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 39.2, pages 414–425 (cited on pages 16, 46).

McGookin, D., Brewster, S. and Jiang, W. W. (2008). Investigating touchscreen accessibility for people with visual impairments. In *Proceedings of NordiCHI*, pages 298–307 (cited on page 19).

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press (cited on page 17).

Mistry, P. and Maes, P. (2010). Mouseless. In *Adjunct Proceedings of UIST (Poster)*, pages 441–442 (cited on page 10).

Mistry, P., Maes, P. and Chang, L. (2009). WUW – Wear Ur World: a wearable gestural interface. In *Extended Abstracts of CHI (Poster)*, pages 4111–4116 (cited on pages 12, 13).

Moeslund, T. B., Hilton, A. and Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding* 104.2-3, pages 90–126 (cited on page 16).

Morelli, T. and Folmer, E. (2012). TWuiST: a discrete tactile-proprioceptive display for eye and ear free output on mobile devices. In *Proceedings of HAPTICS*, pages 443–450 (cited on page 10).

Morris, C. D., Bransford, J. D. and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior* 16.5, pages 519–533 (cited on page 20).

Morris, D., Saponas, T. S. and Tan, D. (2010). Emerging input technologies for always-available mobile interaction. *Foundations and Trends in Human–Computer Interaction* 4.4, pages 245–316 (cited on page 15).

Moscovich, T. and Hughes, J. (2004). Navigating documents with the virtual scroll ring. In *Proceedings of UIST*, pages 57–60.

Mulder, A. (1996). *Hand gestures for HCI*. Technical Report. Simon Fraser University. URL: `http://www.xspasm.com/x/sfu/vmi/HCI-gestures.htm#HCI` (cited on page 17).

Nakatsuma, K., Shinoda, H., Makino, Y., Sato, K. and Maeno, T. (2011). Touch interface on back of the hand. In *Proceedings of SIGGRAPH Emerging Technologies*, article no. 19 (cited on pages 13, 15).

Ni, T. and Baudisch, P. (2009). Disappearing mobile devices. In *Proceedings of UIST*, pages 101–110 (cited on pages 1, 11, 24, 25, 26, 28, 29).

Oakley, I. and Park, J. (2009). Motion marking menus: an eyes-free approach to motion input for handheld devices. *International Journal of Human-Computer Studies* 67.6, pages 515–532 (cited on page 14).

Oikonomidis, I., Kyriazis, N. and Argyros, A. A. (2012). Tracking the articulated motion of two strongly interacting hands. In *Proceedings of CVPR*, pages 1862–1869 (cited on page 16).

Olberding, S., Yeo, K. P., Nanayakkara, S. and Steimle, J. (2013). AugmentedForearm: exploring the design space of a display-enhanced forearm. In *Proceedings of Augmented Human*, pages 9–12 (cited on page 12).

Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J. and Gançalves, D. (2011). Blind people and mobile touch-based text-entry: acknowledging the need for different flavors. In *Proceedings of ASSETS*, pages 179–186 (cited on page 19).

Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of CHI*, pages 919–928 (cited on page 14).

Park, K. S., Hong, G. B. and Lee, S. (2012). Fatigue problems in remote pointing and the use of an upper-arm support. *International Journal of Industrial Ergonomics* 42.3, pages 293–303 (cited on page 48).

Perkins, D. N. and Salomon, G. (1992). Transfer of learning. In *International Encyclopedia of Education*. 2nd edition (cited on page 20).

Perugini, S., Anderson, T. J. and Moroney, W. F. (2007). A study of out-of-turn interaction in menu-based, IVR, voicemail systems. In *Proceedings of CHI*, pages 961–970 (cited on page 61).

Pietroszek, K. and Lank, E. (2012). Clicking blindly: using spatial correspondence to select targets in multi-device environments. In *Proceedings of MobileHCI*, pages 331–334 (cited on page 9).

Pirhonen, A., Brewster, S. and Holguin, C. (2002). Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of CHI*, pages 291–298 (cited on page 19).

Pulli, P. and Antoniac, P. (2004). User Interface. Patent 6771294 (US) (cited on page 12).

Raghunath, M. T. and Narayanaswami, C. (2002). User interfaces for applications on a wrist watch. *Personal and Ubiquitous Computing* 6.1, pages 17–30 (cited on page 11).

Rahman, M., Gustafson, S., Irani, P. and Subramanian, S. (2009). Tilt techniques: investigating the dexterity of wrist-based input. In *Proceedings of CHI*, pages 1943–1952 (cited on page 15).

Rekimoto, J. (2001). GestureWrist and GesturePad: unobtrusive wearable interaction devices. In *Proceedings of ISWC*, pages 21–27 (cited on pages 11, 16, 58).

Rizzolatti, G., Fadiga, L., Fogassi, L. and Gallese, V. (1997). The space around us. *Science* 277, pages 190–191 (cited on page 13).

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D. and Dantzich, M. V. (1998). Data Mountain: using spatial memory for document management. In *Proceedings of UIST*, pages 153–162 (cited on page 7).

Rosenberg, R. and Slater, M. (1999). The chording glove: a glove-based text input device. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 29.2, pages 186–191 (cited on page 15).

Saponas, T. S., Harrison, C. and Benko, H. (2011). PocketTouch: through-fabric capacitive touch input. In *Proceedings of UIST*, pages 303–308 (cited on page 14).

Sato, M., Poupyrev, I. and Harrison, C. (2012). Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of CHI*, pages 483–492 (cited on page 15).

Shneiderman, B. (1983). Direct manipulation: a step beyond programming languages. *IEEE Computer* 16, pages 57–69 (cited on page 17).

Shoemaker, G., Tsukitani, T., Kitamura, Y. and Booth, K. S. (2010). Body-centric interaction techniques for very large wall displays. In *Proceedings of NordiCHI*, pages 463–472 (cited on page 12).

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A. (2011). Real-time human pose recognition in parts from single depth images. In *Proceedings of CVPR*, pages 1297–1304 (cited on page 18).

Sibert, R. W. L. J. L. and Hahn, J. K. (1999). Hand-Held Windows: towards effective 2D interaction in immersive virtual environments. In *Proceedings of IEEE VR*, pages 205–212 (cited on page 13).

Starner, T., Auxier, J., Ashbrook, D. and Gandy, M. (2000). The Gesture Pendant: a self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Proceedings of ISWC*, pages 87–94 (cited on pages 1, 18, 45, 58).

Starner, T., Weaver, J. and Pentland, A. (1997). A wearable computer based american sign language recognizer. In *Proceedings of ISWC*, pages 130–137 (cited on page 18).

Steins, C., Gustafson, S., Holz, C. and Baudisch, P. (2013). Imaginary Devices: gesture-based interaction mimicking traditional input devices. In *Proceedings of MobileHCI*, pages 123–126 (cited on page 107).

Strachan, S., Murray-Smith, R. and O'Modhrain, S. (2007). BodySpace: inferring body pose for natural control of a music player. In *Extended Abstracts of CHI (Poster)*, pages 2001–2006 (cited on page 12).

Sturm, I., Schiewe, M., Köhlmann, W. and Jürgensen, H. (2009). Communicating through gestures without visual feedback. In *Proceedings of Pervasive Technologies Related to Assistive Environments*, article no. 15 (cited on page 11).

Sturman, D. J. and Zeltzer, D. (1994). A survey of glove-based input. *Computer Graphics and Applications* 14.1, pages 30–39 (cited on page 18).

Sturman, D. J. and Zeltzer, D. (1993). A design method for "whole-hand" human-computer interaction. *Transactions on Information Systems* 11.3, pages 219–238 (cited on pages 17, 18).

Tamaki, E., Miyaki, T. and Rekimoto, J. (2009). Brainy Hand: an ear-worn hand gesture interaction device. In *Extended Abstracts of CHI (Poster)*, pages 4255–4260 (cited on pages 12, 13, 16).

Tan, D. S., Stefanucci, J. K., Proffitt, D. R. and Pausch, R. (2002). Kinesthetic cues aid spatial memory. In *Proceedings of CHI*, pages 806–807 (cited on page 8).

Ukita, N. and Kidode, M. (2004). Wearable virtual tablet: fingertip drawing on a portable plane-object using an active-infrared camera. In *Proceedings of IUI*, pages 169–176 (cited on page 14).

Vallbo, A. B. and Johansson, R. S. (1978). The tactile sensory innervation of the glabrous skin of the human hand. In *Active Touch, the Mechanism of Recognition of Objects by Manipulation*. Edited by G. Gordon. Pergamon Press Ltd., pages 29–54 (cited on pages 21, 99, 103).

Vanderheiden, G. C. (1996). Use of audio-haptic interface techniques to allow nonvisual access to touchscreen appliances. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*. Poster, page 1266 (cited on pages 19, 84).

Vardy, A., Robinson, J. and Cheng, L. T. (1999). The WristCam as input device. In *Proceedings of ISWC*, pages 199–202 (cited on page 16).

Vogel, D. and Baudisch, P. (2007). Shift: a technique for operating pen-based interfaces using touch. In *Proceedings of CHI*, pages 657–666 (cited on pages 80, 82).

Voisin, J., Lamarre, Y. and Chapman, C. E. (2002). Haptic discrimination of object shape in humans: contribution of cutaneous and proprioceptive inputs. *Experimental Brain Research* 145, pages 251–260 (cited on page 21).

Wachs, J. P., Kölsch, M., Stern, H. and Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM* 54.2, pages 60–71 (cited on page 16).

Wagner, J., Nancel, M., Gustafson, S., Huot, S. and Mackay, W. E. (2013). A body-centric design space for multi-surface interaction. In *Proceedings of CHI*, pages 1299–1308 (cited on pages 13, 107).

Wallet, G., Sauzéon, H., Rodrigues, J. and N'Kaoua, B. (2009). Transfer of spatial knowledge from a virtual environment to reality: impact of route complexity and subject's strategy on the exploration mode. *Journal of Virtual Reality and Broadcasting* 6.4, pages 572–574 (cited on page 20).

Wang, F. and Ren, X. (2009). Empirical evaluation for finger input properties in multi-touch interaction. In *Proceedings of CHI*, pages 1063–1072 (cited on page 80).

Wang, Q., Hsieh, T. and Paepcke, A. (2009). Piles across space: breaking the real-estate barrier on small-display devices. *International Journal of Human-Computer Studies* 67.4, pages 349–365 (cited on page 9).

Wang, R. Y. and Popović, J. (2009). Real-time hand-tracking with a color glove. In *Proceedings of SIGGRAPH*, article no. 63 (cited on page 17).

Wang, Y. and MacKenzie, C. L. (2000). The role of contextual haptic and visual constraints on object manipulation in virtual environments. In *Proceedings of CHI*, pages 532–539 (cited on page 13).

Wigdor, D. and Wixon, D. (2010). *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann.

Wikipedia (2013). *Guidonian Hand*. URL: `http://en.wikipedia.org/wiki/Guidonian_hand` (visited on 04/15/2013) (cited on page 13).

Willis, K. D. D., Lin, J., Mitani, J. and Igarashi, T. (2010). Spatial Sketch: bridging between movement & fabrication. In *Proceedings of TEI*, pages 5–12 (cited on page 10).

Wilson, A. D. (2010). Using a depth camera as a touch sensor. In *Proceedings of ITS*, pages 69–72 (cited on pages 17, 58).

Wilson, A. D. (2006). Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proceedings of UIST*, pages 255–258 (cited on pages 44, 47).

Wilson, A. D. and Oliver, N. (2003). GWindows: robust stereo vision for gesture-based control of windows. In *Proceedings of ICMI*, pages 211–218 (cited on page 18).

Witt, H., Lawo, M. and Drugge, M. (2008). Visual feedback and different frames of reference: the impact on gesture interaction techniques for wearable computing. In *Proceedings of MobileHCI*, pages 293–300 (cited on page 10).

Wolf, K., Naumann, A., Rohs, M. and Müller, J. (2011). A taxonomy of microinteractions: defining microgestures based on ergonomic and scenario-dependent requirements. In *Proceedings of INTERACT*, pages 559–575 (cited on page 14).

Wolf, K., Schleicher, R., Kratz, S. and Rohs, M. (2013). Tickle: a surface-independent interaction technique for grasp interfaces. In *Proceedings of TEI*, pages 185–192 (cited on page 14).

Yang, X.-D., Grossman, T., Wigdor, D. and Fitzmaurice, G. (2012). Magic Finger: always-available input through finger instrumentation. In *Proceedings of UIST*, pages 147–156 (cited on page 15).

Yee, K.-P. (2003). Peephole Displays: pen interaction on spatially aware handheld computers. In *Proceedings of CHI*, pages 225–234 (cited on page 9).

Yin, M. and Zhai, S. (2006). The benefits of augmenting telephone voice menu navigation with visual browsing and search. In *Proceedings of CHI*, pages 319–328 (cited on page 61).

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R. and Baudisch, P. (2007). earPod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of CHI*, pages 1395–1404 (cited on page 19).

Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S. and Harvill, Y. (1987). A hand gesture interface device. In *Proceedings of CHI*, pages 189–192 (cited on page 18).

COLOPHON

This dissertation was typeset with *Palatino* for text and math, *Iwona* for headings and *Euler* for chapter numbers. The style is based on the ArsClassica customization of the ClassicThesis LaTeX template.

Review version produced June 24, 2013.

This version produced November 24, 2013, differing only by small typographical and grammatical fixes.