Swansea University
Prifysgol Abertawe

Eyes-Off Physically Grounded Mobile Interaction

Simon Robinson

A thesis submitted in partial fulfilment for the
degree of Doctor of Philosophy

in the
Future Interaction Technology Laboratory
Department of Computer Science
College of Science
Swansea University

October 2012

Abstract

This thesis explores the possibilities, challenges and future scope for eyes-off, physically grounded mobile interaction. We argue that for interactions with digital content in physical spaces, our focus should not be constantly and solely on the device we are using, but fused with an experience of the places themselves, and the people who inhabit them. Through the design, development and evaluation of a series of novel prototypes we show the benefits of a more eyes-off mobile interaction style. Consequently, we are able to outline several important design recommendations for future devices in this area.

The four key contributing chapters of this thesis each investigate separate elements within this design space. We begin by evaluating the need for screen-primary feedback during content discovery, showing how a more exploratory experience can be supported via a less-visual interaction style. We then demonstrate how tactile feedback can improve the experience and the accuracy of the approach. In our novel tactile hierarchy design we add a further layer of haptic interaction, and show how people can be supported in finding and filtering content types, eyes-off.

We then turn to explore interactions that shape the ways people interact with a physical space. Our novel group and solo navigation prototypes use haptic feedback for a new approach to pedestrian navigation. We demonstrate how variations in this feedback can support exploration, giving users autonomy in their navigation behaviour, but with an underlying reassurance that they will reach the goal.

Our final contributing chapter turns to consider how these advanced interactions might be provided for people who do not have the expensive mobile devices that are usually required. We extend an existing telephone-based information service to support remote back-of-device inputs on low-end mobiles. We conclude by establishing the current boundaries of these techniques, and suggesting where their usage could lead in the future.

Keywords: Eyes-off, physically grounded, location-based, mobile interaction, haptics.

Acknowledgements

I would like to thank the many people and organisations who have supported me throughout this research. This thesis would never have come together without the friendship, support and insights from so many others.

First and foremost, thanks to Matt Jones, my supervisor, mentor and friend before, during and hopefully after this PhD. Matt's insights and guidance are intertwined throughout this thesis, and without his guidance I would never have got this far. Thanks also to my second supervisor, Parisa Eslambolchilar, for her help and advice in the early stages of this work, and to Keith Cheverst and Mark Dunlop, who examined this research during my viva.

My friends and colleagues in the FIT Lab – particularly Jen, Patrick, Tom and Ben, have helped both during this work and (especially) in the much-needed breaks and distractions – thank you all. Elina, thank you so much for doing my real job while I concentrated on other things for a while!

The financial support of EPSRC (from projects EP/E042171/1, EP/H042857/1 and EP/G003971/1), IBM Research India and Swansea University's Computer Science Department has funded this research throughout, and for this I am particularly grateful. Collaborators and colleagues on the various projects related to this work have helped immensely; in particular I would like to thank John, Rod, Stephen, Craig, Amit, Nitendra, Anupam, Kristen and David.

This thesis is dedicated to my family and friends, without whose love and support I probably would not have even started.

Contents

Chapter One

Introduction

The first time seeing any prediction of the future of technology is often exciting and inspiring, sometimes even quite magical. When it comes to visions about mobile devices, the situation is no different. But the future of mainstream, commercial mobile interaction is also remarkably single-minded, seemingly set to revolve around touching, swiping; at most, speaking to or bending a thin, screen-primary mobile device. While the modern mobile is incredibly powerful, interaction is very often little more than a simulation of physical buttons and sliders: touchscreen "pictures under glass" [153]. Furthermore, when considering the vast number of devices available today, it is clear that the fundamental ways we interact with mobiles have changed surprisingly little since the mobile phone was first envisaged. Early mobiles used buttons and small screens, while today's devices have simply merged keypad and screen into one. Using these devices requires looking at their screens, so as a result, people wander around, heads down, stroking the smooth glassy surfaces of their devices, captivated by their personal digital world.

In this thesis we step back to consider how to design for mobile interactions that help us explore and interact with our surroundings in a non screen dominant way. The world around us is full of fantastic things to see, explore and share – we suggest that it is a waste to spend our time looking at a screen to stay connected, located and up-to-date. The prototypes presented in this thesis aim at allowing exploration, letting people point, feel or move to interact—eyes-off—with their mobiles. Our goal, then, is to, through a series of prototypes and evaluations, ground the everyday interactions we have with our mobile devices more in the physical world we live in.

1.1 Physically grounded mobile interaction

Currently, the everyday interactions we have with modern mobile devices are focused on the device itself, rather than on the people and places directly around us. However, the huge popularity of services such as Facebook or Foursquare for sharing what we are doing with our lives demonstrates an underlying desire to see and share experiences. At present, to experience such digital content associated with real places we must direct our attention toward the technology framework that was designed to help us, rather than the physical world we live in. We argue that while this allows us to maintain general awareness of our surroundings—we rarely walk into things, for example—it can lead to losing touch with the world around us, and missing experiences.

Indeed, recently, it seems that this sentiment is becoming more widespread. Turkle [150], for example, suggests that we are becoming "alone together," living lives that are physically present but digitally absent: we bring our connectivity and screen-received updates with us wherever we go, never truly experiencing other people or places. Victor [153] denounces future devices that are based on touchscreens rather than tactile richness, and Jones [72] argues for a return to more personal, imaginative or even extravagant experiences of places, rather than the benign, digitally-filtered world we interact with currently.

Ubiquitous awareness has clearly become a common and everyday desire. To see how this has come about, however, we should look back at the evolution of mobile ubiquitous computing since its foundation in Weiser's original ubicomp vision [158]. The central argument of the original ubicomp article was that technology should be designed to fade into the background. Rather than directing focus to a screen, or virtualising the physical world, Weiser argued, we should design for the opposite: 'embodied virtuality.' Embodied virtuality blurs the boundaries between the physical and the digital, allowing 'embodied interaction' with elements in the physical and social worlds around us [38]. Essentially, Weiser argued, by bringing computers more naturally into our lives, and having them enhance the world around us, adapting their behaviour based upon where we are and the task at hand, we could eventually allow them to disappear from our conscious thought – reaping their benefits without the need to concentrate on having to use them.

In many respects this vision has come to pass – and it is primarily due to the introduction of the modern smartphone. The first digital (GSM) mobile phone call was made in the same year as Weiser's article. By the end of 2001 there were 0.9 billion mobile phone subscriptions worldwide [66] – a 2011 UN estimate puts the number at 5.9 billion[1] [67]. Early mobiles were simply for phone calls; later versions gradually adapted to become multifunction smartphones. This rapid rise of increasingly mobile devices was coupled simultaneously with huge advances in computing power, and the emergence of wireless connectivity anywhere. More importantly, thanks to compact, low-power GPS receivers, modern mobiles have become capable of constant location awareness. As a result, there has been a strong and growing research and commercial interest in bringing digital content to our experience of physical places.

Initial location-aware mobile devices, such as the ParcTab [155], showed how personal technology could be augmented with location- and context-awareness by equipping buildings with sensor networks. Shortly after, tourist systems such as Cyberguide [92] began to bridge the gap between indoor and outdoor location-aware systems, providing geolocated tourist information based upon the position and orientation of the user. Later systems, such as GUIDE [25], allowed city visitors to create context-aware tours and view maps or geolocated content, focusing closely on the experiences and interests of individual users. More recently, Rukzio [122] sought to define a framework for physical mobile interaction, concentrating on augmenting physical objects with tags or other markers that represented or highlighted digital content. In the past few years, augmented reality browsers such as Layar [85] (building upon the ideas of early tag-based systems such as NaviCam [119]) have brought real-time digital augmentation of the physical world to the screen of any modern handset.

Just like in the numerous early ubiquitous computing scenarios that many researchers and designers have outlined, these mobile devices and services have brought new ways to digitise our lives. These new interactions are extensive, ranging from augmenting live views of the real world with geolocated objects to providing realtime updates about the places we live in and the things we see. Yet while these technological innovations have encouraged and brought about great progress, one key design element has remained at the forefront throughout this transformation: interaction with these devices focuses on the visual modality, and using a device requires looking

---

[1] The total number of mobile phone subscriptions in many regions is as high as 2.1 per capita; actual worldwide mobile device penetration is estimated at approximately 60% in 2011 [21, 149].

at its screen. Weiser's original article stressed the invisibility of computers – not in the physical sense, but philosophically: mobiles should be ready-to-hand rather than just present-at-hand [162]. The aim was for mobiles and computers that did not distract the user from their main task. However, along the way, the dream of devices that step aside to let us continue our normal lives has turned into mobiles that capture our visual attention. But why is this the case?

Put simply, using a screen is quick and easy for many of the most common mobile interactions. While other modalities—haptics or audio, for example—certainly enhance the experience, they do not necessarily speed it up. Speed is perhaps the most common measure of how well people are able to achieve a task[2] and is, of course, a particularly important measure of the usability of many systems. But while greater speed is certainly better for many tasks, especially those that are necessarily screen-primary, we argue that a total focus on time-dependency is not always appropriate for tasks such as physical mobile interaction, browsing, or exploration.

Interaction with geolocated content is a perfect example of this: while the current focus on quick screen-based interaction offers plenty of support for viewing or browsing digital content in situ, there is little support for using this content to complement the *physical* experience of a place. Existing commercial approaches, such as augmenting a camera view with digital overlays, merely place a digital lens—a barrier—in front of our physical view, dividing our attention between the real and the virtual.

This need to divide attention seems likely to limit the much longed for (at least in the research community; e.g., [38, 68, 82, 119, 123]) vision of a fusion of the physical and digital worlds. We argue that, in many scenarios, interactions that are less focused on the screen show the potential to improve mobile experiences, and could allow people to withdraw from the grip of a mobile device's display. While many people clearly desire to spend their entire lives connected, it seems that in some cases there are better interaction methods than screen-focused information streams. Indeed, while screen interaction is clearly appropriate for many—perhaps even most—tasks, the need for such interaction is less apparent when considering physically-grounded situations such as location- or context-aware browsing (rather than retrieval or storage), or pedestrian navigation. Instead, we argue that we should be moving toward interacting with the world, supported, but not controlled, by our mobiles.

---

[2] See the profusion of Fitts' law HCI research since MacKenzie's early paper [97], for example, or the abundance of work, including our own, that uses task completion time as a measure.

1.2 Eyes-off interaction

Rather than constantly looking down at a device's screen, in this thesis we envisage people looking at the world around them, fusing their digital experience with the real-life view of the physical place they are in. Further, we imagine people using mobiles to interact directly with the features in the environment. Instead of recreating the physical world digitally, we would prefer to allow people to use mobile devices as tools to interact with real elements in the world around them. In this thesis, we define this class of interaction as 'eyes-off.'

Interactions that do not require the user to look at the screen have previously been studied in the similar forms of 'heads-up' or 'eyes-free' interaction ([15, 92], for example). However, much of the previous research has proposed entirely eyes-free interaction—the use of a device without ever needing to look at it—or relies on users wearing glasses or headphones to interact. Here we take a more pragmatic approach, acknowledging that in many situations the visual modality is indeed more appropriate. We focus our efforts on re-imagining several situations where interaction that is not screen-primary could help complement existing approaches, allowing people to use mobiles eyes-off. We argue for eyes-off interaction as a divergence from more typical eyes-free or heads-up styles – while clearly similar, our aim is to empower users to directly interact with the people and places around them, rather than a digital representation.

While a more futuristic focus on heads-up only (e.g., entirely speech- or glance-controlled) interaction could indeed lead to complete immersion in the physical world rather than the digital, our approach is to use interaction styles and hardware components that are currently available on standard commercial mobile platforms, and are likely to continue to be part of mobile interaction for some time. Our prototypes primarily use haptic feedback and pointing-based interaction via inertial sensors.

Rekimoto [117] pioneered the use of accelerometers for orientation estimation, and Hinckley et al. [60] subsequently demonstrated the use of inertial sensors for control of a device's screen orientation, paving the way for the use of on-device sensors to react to external conditions. In the decade since, this has become the norm in many situations. Brewster et al. [15] motivated using these types of newly-capable devices with eyes-free interaction techniques using the example of 2D and 3D audio combined with head or finger gestures. Since then, many authors have described novel techniques, including olfaction [96], gestures [144], electromyography [31], shape changes [59], clapping [77], pressure [14], rhythm [152] or weight shifting [58], to name but a few.

Many of these interaction styles can be seen as a step towards *tangible* interaction with technology, a concept originally proposed by Ishii and Ullmer [68]. This influential article presented a vision of 'tangible bits,' with the aim being *"to bridge the gaps between both cyberspace and the physical environment, as well as the foreground and background of human activities"* [68]. The goal was to make the world, rather than the computer, the interface. However, Ishii and Ullmer's aim was not only to allow people to look at the world around them, by freeing their *visual* attention; users should also not have to focus their *cognitive* attention on the device. As Cheverst et al. [27] argued, context-aware devices should automatically and predictably react to environmental triggers to push (or let the user pull) relevant information, without unnecessarily distracting a person's attention from their primary task.

Researchers and news outlets have complained repeatedly about the issue of attention in screen-primary, heads-down interaction with mobile devices. The most seemingly widespread example is that of pedestrians walking into street furniture (or other people) when looking at their mobiles instead of their environment (see [151] for a brief review of such studies and articles). However, when investigating this issue further, it becomes clear that the problems that stem from looking down at a device are not as obvious and clear-cut as they might seem. Indeed, it is remarkably easy to verify this – simply walk down any crowded street and it is actually quite difficult to walk into people, even when they are using their mobiles heads-down.

When looking deeper, these arguments around colliding pedestrians turn out to be quite facile, glossing over the underlying complexities of human behaviour. Although people are indeed concentrating on their personal screens, this concentration is not entirely at the expense of other senses. While pedestrians walking heads-down are clearly distracted from the world around them, often having to divide attention between a screen and surroundings [5], they are not completely unaware of what is happening in their nearby environment.

This background processing—situational awareness—of other environmental factors has long been studied by human factors researchers [42]. Simons and Chabris

[136] provided evidence of the potential effects of concentrating on a particular task—inattentional blindness—in a study since replicated focusing particularly on the use of mobile phones [65]. The authors note that while people may miss background, less important events while concentrating on objects at the centre of interest, these events can still have implicit effects on their behaviour. In this thesis, we are partly motivated by the seemingly insignificant background events that are part of the excitement and mystery of life – the serendipitous discoveries and chance meetings that enrich our experiences.

1.3 Our focus

The research in this thesis aims to give people the chance to actually see the less important things as well as those that have their immediate focus. Eyes-off designs, then, are situated, and involve interaction (either explicit or implicit) with the people and places around the user, rather than simply their digital representations on a mobile device. As demonstrated here, this situated interaction may occur either via the system's knowledge of a user's physical location, or, as in our final chapter, via people interacting eyes-off with a service in a specific place for a particular purpose. We investigate a wide range of situated interactions in this thesis, including explicit pointing to places or objects eyes-off to discover information about them; eyes-off navigation where feedback is minimal and not necessarily related to the current location of the user; and, distal eyes off interaction where the user's location is implicit and not known by the device – situated interaction here is via the user's queries to a remote system.

In our definition, eyes-off designs use a mobile device as a conduit for interaction, rather than as a foreground artefact. Unlike, for example, Strachan et al. [144], who used explicit positioning of a device around the user's body to control a music player, we focus on using the device to help the user explore a physically-grounded information space. Our systems are intrinsic, egocentric and relative to the user's current frame of view, rather than exocentric, as discussed by Bidwell and Lueg [8]. The interactions supported by eyes-off systems do not attempt to simulate physical objects using tactile feedback, like the designs of Williamson et al. [160] for example. Rather, they rely on the user themselves to interpret the consistent feedback that is given and relate this to their expectations of the system.

While our aims overlap with the definition of 'eyes-free' given by Oakley and Park [105], we do not share their desire for interaction "without vision." Instead, our designs focus on the use of a mobile device in the background, rather than the foreground of interaction: people focus on a system or service as a whole, rather than the device itself. This is most clear in our navigation work, where people use the prototype as a tool to discover a goal, rather than needing to look at the device itself to interact, or in our low-end back-of-device interaction, where people control a remote service by tapping on the back of a phone. This does not mean that there is no need for a visual, screen-primary interface, however – we aim for a fusion of both eyes-off and screen-based interactions as in, for example, our pointing-based haptic content filtering system, which allows viewing in situ.

Our eyes-off paradigm supports both exploratory and focused interactions. This flexibility is not achieved via background coercion, however, as in nudge-based interaction [147]. Rather than aiming to encourage people into interacting a particular way or performing a specific behaviour via some level of artificial freedom, we design for exploration and adaptability in interaction. In order to do this, our designs and prototypes have the common aim of letting people focus on the world around them rather than their mobile, supporting interaction with eyes off the device rather than eyes-on looking at a screen.

Throughout this thesis we have three general research questions that help to explore eyes-off interaction. Our primary focus is on exploring engagement, and the extent to which eyes-off designs allow people to interact with their surroundings rather than their device. To measure this, participants are observed during user studies, and interviewed in a semi-structured manner after using our prototypes. In several cases we also use quantitative measures of engagement and cognitive load, via TLX [56] and PPWS [111].

Our second research focus is on comparing the differences in user behaviour and interaction styles between traditional, screen-primary systems, and our eyes-off designs. These differences are measured, where appropriate, by constructing equivalent screen-based alternatives to our eyes-off prototypes, and comparing these in efficiency and interaction time required. We also measure support for engagement with surroundings by requiring participants to navigate, for example, around an obstacle course, or to a specific destination.

The final focus of this thesis is on exploring the viability of eyes-off interaction styles as a primary method for achieving mobile tasks. In order to investigate this we measure the time taken to interact with our systems, and the cognitive load imposed (via TLX and PPWS). This question changes throughout this thesis, however, as our ideas and designs shift. We move from purely time-dependent discovery of geolocated content to an exploratory view of navigation, and then to focused interactions with no location-based hardware required. This shift is revisited in our concluding chapter, where we discuss the changing motives and questions throughout this research.

At the start of the research for this thesis, location, motion and other sensors were only just becoming widely available in commercial mobile devices. Still, the potential for using these sensors to provide rich gestures and multimodal interactions in situ—using the affordances of the physical world to access digital elements—seemed great. More recently, such sensors have become far more widespread, but the main use of these is for augmenting or supporting the display of a device. While researchers continue to develop and evaluate novel mobile interactions that embrace physical affordances, mainstream commercial devices are still focused around the screen.

Perhaps, then, modern mobiles will not change the way we interact with the world? The current situation is in fact surprisingly reminiscent of other early technology predictions, such as the 'paperless office' [129] where today, despite huge research advances, paper usage is still increasing [148]. Just like the mainstream screen-primary focus of mobile interaction, this situation does not look set to change in the immediate future. In this thesis, we aim to address this by remembering, as Jones and Marsden [71, p. 52] argue, that we are innately ecological, not technological. Just like recent work that is starting to bring the affordances of paper to digital reading devices [110], we attempt to remove the focus from the screen to the physical things that we interact with. Importantly, though, unlike e-readers, our focus on interaction styles and modalities that are available on current platforms provides a way to potentially fast forward the adoption of eyes-off interaction in the mobiles of the future.

1.4 Overview and contributions

This research explores the possibilities, challenges and future scope for mobile interactions that can support and enable a fusion of physical and digital spaces via eyes-off interactions. The four key contributing chapters in this thesis each investigate separate elements within this design space, embodying these in novel mobile prototypes that advance the state-of-the-art. These prototypes, and the results of their evaluation, help us to propose a set of design recommendations for future systems in this area.

In *Discovering* we investigate the use of both screen-primary and non-visual modalities for interacting with geolocated digital content in physical environments. We show how eyes-off feedback can allow more exploratory discovery behaviour and, in the second half of this chapter, how the addition of haptic feedback can improve the experience.

Turning to *Displaying*, we examine techniques for allowing users to browse the content that has been discovered. We show how a two-level haptic feedback hierarchy can allow users to find and filter particular content types, and demonstrate its usability and accuracy in two studies. We then turn to consider how the locations of other people in the nearby locality can be displayed to each other, via a central group meeting point. The haptic prototype we evaluate allows people to easily follow a vibrotactile cue, eyes-off, and rendezvous with other users.

In *Shaping*, we take a broader view of physically-grounded interactions, exploring how people's movements through their wider physical environment can be shaped by the device they use. We demonstrate several designs for a new approach to pedestrian navigation, showing how the turn-by-turn methods currently used may not be particularly appropriate, or even necessary. We show how the use of haptic feedback, as a function both of available routes and of social media content, can be used to give people a sense of autonomy in navigation, but with an underlying confidence that they can reach their goal.

The final contributing chapter of this thesis turns to consider mobile interaction for *The Next Billion People*, focusing on how some of the types of advanced eyes-off mobile interactions we have explored in our first three chapters might be supported for the many millions of mobile phone users who do not have—and may never have—access to the sorts of devices we have used. In many developing regions there is an increasing focus on audio as an interaction channel[3] in order to meet the local challenges of textual literacy, cost and lack of data connections. We show in this chapter how a

---

[3] And, indeed, in more developed regions – see the recently released and much publicised *Siri* voice recognition feature on Apple's iPhone.

subset of potential eyes-off interaction methods can be supported remotely, by creating a server-based audio gesture recogniser that allows users of an audio-based voice information service to tap on the back of their devices for input.

As the capabilities of mobile devices continue to increase, the interaction techniques described in this thesis will become increasingly achievable using standard devices – indeed, many of our designs could now arguably be implemented as standard smartphone applications. Consequently, we also aim to develop and support an understanding of eyes-off, physically grounded interactions that can be applied to future mobile devices. Each chapter of this thesis closes, then, with a set of several design recommendations for this class of devices. Finally, we conclude by establishing the current boundaries of these techniques, and suggesting where their use could lead in the future.

1.4.1 The author's contribution

The vast majority of this research was undertaken by the thesis author. For two elements, however, there were significant collaborations with colleagues at external institutions. In particular, approximately 90 % of the programming for the group rendezvous prototype discussed in Section 4.5.1 (Pages 103 to 105) was undertaken by John Williamson (University of Glasgow). For this work, the thesis author's role was to refine the prototype, and to design, manage and analyse the results of the user study undertaken to evaluate the concept. A similar collaborative approach was taken for user studies of our final prototype, where the thesis author designed, planned and developed the prototype, designed user studies and analysed their findings, but the studies themselves (see Sections 6.2.2 and 6.3.1) were managed by colleagues at IBM Research India.

Much of the research within this thesis has been published in co-authored peer-reviewed international conference or journal papers. One of these publications (see [P5]) received the *best paper* award at MobileHCI 2009. Our paper at MobileHCI 2010 ([P7]) was nominated for the same award in the following year, and its concepts were also covered by several print, TV, radio and online media channels.

The publication details and abstracts of those papers that are a major part of this work are reproduced in Appendix A for ease of reference.

CHAPTER TWO

Background

In this chapter we review key concepts and previous related research in order to situate our work. Our goal in this thesis is to ground the everyday interactions we have with our mobile devices more in the physical world we live in, separating our eyes-off approaches into three key themes: *discovering*, *displaying* and *shaping*.

We begin, then, by reviewing research into location-aware interaction with physical objects, surveying previous work from early geolocated systems to the more recent point-to-select geo-querying. We then examine eyes-off interaction and related areas, discussing broadly both previous uses of inertial sensing and the various modalities, such as haptic or audio feedback, that could be or have previously been used to provide such interactions in ways that are less focused on the screen of a mobile device.

In Section 2.3 we turn specifically to the *discovering* and *displaying* elements of this thesis, discussing how in situ browsing and retrieval of geolocated content has previously been approached. Section 2.4 reviews previous research in *shaping* the ways people interact with geolocated content, covering themes from location annotation to less restrictive pedestrian navigation. We conclude with an overview of the issues we have highlighted in this area of mobile interaction, summarising how the research in this thesis addresses and relates to previous work.

2.1 Physically-grounded location-aware interaction

Today's location-aware systems build on a foundation of constant location awareness, combined with knowledge of the user's activities and surroundings, often aiming to provide precise context awareness. These devices and services have been developed over a prolonged period, with many research and technological innovations bringing the state-of-the-art to where it is today.

Coarse location awareness has been available for millennia, in countless forms. For self positioning, methods include dead reckoning, celestial or lunar navigation (improved with tools such as sextants and chronometers), and maps and compasses for land travel. For tracking the location of others, token-based systems used for tracking railway locomotives on single-track lines, and paper- or key-based routines for access to building zones both continue to be used. Precise, automatic mobile location awareness has not been possible until relatively recently, however, with the advent of large-scale widely-available technologies such as GPS, and the more compact modern mobile device.

One of the earliest location aware mobile platforms was the Active Badge system [154], which extended simple building access badges to provide tracking of users' positions. Rather than basic zoning of people's locations by recording rooms users had entered or left, the system used electronic badges which automatically transmitted pulses every 15 s. These pulses were detected by a sensor network installed around the building, and could be placed within a map of the area. This early system was primarily used for administration tasks, such as redirecting landline phone calls to the appropriate location, but was subsequently extended and used as part of the much wider reaching ParcTab.

The ParcTab [155] was a compact, wireless, mobile communication device, building upon the Active Badge sensor network for its location awareness. The system consisted of a touchscreen PDA-like device, equipped with infrared communication between the device and a network of transceivers that provided room-level location resolution. The system was used for sending and receiving data (such as email), and other tasks such as pointing, annotation and group voting, but was also perhaps the first truly context aware platform.

A key part of the ParcTab design was to take into account factors including the user's location, the presence of other devices or people, time of day, and other nearby networked peripherals. This knowledge and incorporation of location, people and resources in the surrounding area has come to be defined as context-aware computing. Context-aware computing was first defined by Schilit et al. [125, 126], who describe it as *"the ability of a mobile user's applications to discover and react to changes in the context in which they are situated"* [126]. While a simplistic view could see context awareness as no different to location awareness, the key defining feature is that rather than simply sensing location, context aware devices are able to also *react* accordingly.

Schilit [126] defined a three-step process through which context aware devices or applications carry out their actions: discovery, in which devices learn about resources and their characteristics; selection, during which the system automatically selects resources based upon their attributes; and use, which may be automatic or manual. The discovery and selection of nearby resources are particularly interesting in the context of this thesis, and Schilit et al. [125] discuss several methods for automatically choosing and selecting resources. These methods are: proximate selection; automatic contextual reconfiguration; contextual information and commands; and, context-triggered actions.

Of these methods, proximate selection is particularly relevant to our research. Proximate selection is a technique where nearby resources are made easier to select, by defaulting to those in the vicinity of the user's current location. Schilit et al. [125] discuss the different kinds of information—devices, people or places—that might be discoverable using this technique. In this thesis, we investigate each of these categories via various point-to-select prototypes.

While early context-aware systems were initially focused towards office applications, tourist systems such as Cyberguide [92] began to bridge the gap between indoor and truly mobile location aware systems. The Cyberguide project aimed to provide information based on the position and orientation of a tourist holding a mobile device, and was available in both indoor and outdoor versions. Designed as a modular tour guide-like architecture, the outdoor version of the system used GPS positioning to guide visitors around a university campus.

Since the development of the Cyberguide system, mobile devices capable of location-awareness have become near ubiquitous. The sensors and beacons that were necessary to provide positioning data are now present in the form of GPS receivers in off-the-shelf smartphones, allowing the creation of richer mobile interactions with far less, if any, investment in infrastructure. These advances have led researchers to aim for of a fusion of the physical and digital worlds, allowing usage of a context-aware mobile device to fluidly interact with real objects in the world around us.

Some researchers have warned against over-reliance on contextual information, however. Schilit et al.'s early work [125] offered caution regarding the challenges of proximate selection, highlighting the need to take into account the large amount of contextual information that could be displayed, and the bandwidth of the mobile device. Cheverst et al. [23] considered further issues with over-reliance on context-aware systems; namely the trust that users are required to place in the system, and the frustration that can occur when the system is unable to accurately and consistently predict the user's intentions.

In more recent commercial context-aware designs, no doubt inspired in part by earlier research systems, it has become common to use a realtime visual display of the surrounding area – whether in the form of a marked-up map, or a camera display augmented with contextual tags. However, we argue that in some cases these devices are unnecessarily overloading the user with a stream of information that is not required for the task at hand. In our designs we have aimed to ground users in their surroundings, with contextual data that is appropriate, rather than simply augmenting them digitally.

2.1.1 Grounding rather than augmenting

Mark Weiser's original ubicomp article [158] made a strong distinction between the notions of ubiquitous computing and virtual reality. Rather than replicating the outside world digitally, he argued for a second stage in the ubiquity of technology: the *invisibility* of many of the devices we use. In this context, invisibility refers less to the physical appearance of a device, and more to the way it is used, however. Central to this invisibility is that computers are designed to become ready-to-hand—*zuhanden* [57]—tools that we use by concentrating on the task, rather than actively thinking about the tools themselves.

Early methods of augmenting the world (e.g., [119]) aimed to provide this device invisibility by using virtual windows. But visually augmenting the physical world is not what we aim for in this thesis. Rather, our focus is on using the affordances of the physical world to support users in the use of their digital devices. Weiser argued similarly, suggesting that rather than augmenting reality, we should reverse this to design for 'embodied virtuality.'

Ishii and Ullmer [68] presented an early vision of 'tangible bits,' the goal of which was to make the world, rather than the computer, the interface. Dourish [38] argues that tangible and social interaction research areas in HCI can be thought of as two different aspects of 'embodied interaction,' which is itself influenced by phenomenology – the philosophical rejection of the idea that the mind and body are separate. Dourish defines embodiment as *"the common way in which we encounter physical and social reality in the everyday world"* [38, p. 100], and uses this argument to develop and propose guidelines for the next generation of systems supporting embodied interaction.

Other proposals have sought to focus less on tangible or embodied interaction with standard interfaces, and more on how we might take advantage of affordances, and cultural and social norms—such as pointing or turning—to interact more naturally. For example, in 1999, Egenhofer [41] proposed several *Spatial Information Appliances* – worn or handheld devices to access geolocated content in situ. The appliances proposed included a Smart Compass, providing turn-based GPS guidance toward a location, and Geo-Wands, that could help users to identify geographic targets by pointing toward them. This prediction was well before the necessary location, motion and other sensory technologies were integrated into mobile devices; in recent years, with the proliferation of sensory hardware, many of these interactions are now possible.

Egenhofer's predictions were an early indication of the beginning of a new interaction paradigm: physical mobile interaction. More recently, Rukzio has defined physical mobile interaction as *"interactions between a user, a mobile device, and a smart object in the real world"* [122]. Rukzio provides a framework for this class of interaction, focusing on the communication channels between users, mobile devices and smart objects. Rukzio's approach is restricted to visual-primary interaction, however. Fröhlich et al. [47] review mobile spatial interaction, focusing more on how spatial sensing can be used to physically ground location-aware mobile interaction. This is similar to the reality-based interaction of [69], who define a more philosophical framework for post-WIMP[1] interfaces. A key component of their framework is a focus on bringing an awareness of the physical world into the digital interactions we have. Our work follows a similar trajectory, but rather than focusing on modelling digital devices on physical interactions, we have concentrated on lessening the need for screen-primary interactions – our prototypes aim to support *eyes-off* interaction.

---

[1] Windows, Icons, Menus and Pointer.

2.2 Mobile eyes-off interaction

In this thesis we explore the design space around several prototypes that demonstrate eyes-off physically grounded interaction. Our work here is different to previous research in, say, heads-up or eyes-free interaction, however. Unlike heads-up interaction, which is primarily based on augmented reality systems, or the entirely non-screen aims of eyes-free interaction, our aim is not to replace traditional, screen-primary interaction in every scenario, but instead only when necessary and beneficial.

2.2.1 Exploring eyes-off interaction

One of the problems within mobile eyes-free or heads-up research is the lack of any common or universal definitions of exactly what work in these areas aims to achieve. While early heads-up research focused solely on scenarios such as displays for aircraft instruments [157], the transition to mobile devices has led to devices and systems with more loosely defined aims. For example, Rekimoto and Nagao's motivation was to allow hands-free operation of a PDA, by using a headset that did not block the user's normal vision [119], while Soute and Markopoulos [138], applied heads-up interaction to gaming, with a focus on minimal equipment, stimulating imagination and "rich, social interaction." Brewster [13] describes heads-up interaction styles in a broader sense as *"interactions that allow people to get on with their lives whilst using the technology."*

Eyes-free interaction covers a similarly broad and overlapping area. Oakley and Park summarise the fundamental motivation for eyes-free devices as interaction that *"leaves visual attention unoccupied, [so] users are free to perform additional tasks"* [105]. The authors also provide a definition of a typical eyes-free system:

> *"[An] interactive system with which experts can interact confidently in the absence of graphical feedback. The system should be aimed towards the general public, should feature an UI which enables a novice user to pick it up and use it immediately and should not rely on complex recognition technologies."*
>
> Oakley and Park [105]

This definition broadly encompasses most possibilities for eyes-free and eyes-off interaction but, like previous proposals, eventually aims for the complete removal of standard, screen-primary interaction methods, rather than refining to focus on eyes-off interaction only when suitable. While Oakley and Park do leave space for simpler interaction designs that can be used without unnecessary complexity, they are clear that eyes-free interfaces must be operated "without vision."

This aim of removing standard interaction techniques often necessitates using additional devices, such as wristbands or waistbands, or other sensors around the body. While much of the previous eyes-free interaction research has focused on such specialty, customised devices or, alternatively, on heads-up techniques such as visually augmented glasses, in this thesis we target hardware that is now present in commercially-available devices, and will likely be available for some time.[2] Eyes-free interaction designs that focus on interacting without *ever* having to look at the device are, we argue, unrealistic in many situations – while mobile devices are an essential part of most day-to-day lives, it is a significant extra step to adopt wearables or augmented glasses for everyday use, and the benefits are not clear [80].

In this thesis we define eyes-off mobile interaction as diverging from eyes-free interfaces to use physically-grounded interactions to access digital content. But while the eyes-off designs we aim for are clearly often similar to eyes-free designs, we are not proposing the removal of all visual or screen-based interfaces. Instead, by learning from the incomplete and often poor adoption of previous heads-up and eyes-free designs, we aim to address scenarios and contexts where looser, more flexible approaches can help offer a richer user experience. While many researchers have proposed futuristic devices and platforms for the mobiles that lie ahead, our work offers perhaps an intermediate step between the screen-focused designs of the present and the invisible devices of the future.

2.2.2 Beyond heads-up and eyes-free interaction, to *eyes-off*

While early heads-up applications were usually designed for military scenarios [157], more recent mobile heads-up interaction is often based around glasses or augmented camera displays for augmented reality. Rekimoto and Nagao's early work in this area

---

[2] For the prototypes in this thesis we have used separate inertial sensor packs and, initially, external GPS receivers in several designs, to support the types of eyes-off interaction we propose during the transition from featurephone to smartphone. These sensors are now widely present in current mobile devices, however, and the techniques used throughout this thesis are entirely possible on modern mobiles.

used digital overlays of physical objects on a mobile screen to display information about the physical objects in view [119]. Although their system was not fully eyes-free or heads-up when compared to more recent devices, it paved the way for visual augmented reality systems, such as the glasses-based systems described by Starner et al. [140], and popular commercial camera-based browsers (Layar, for example [85]).

Starner's glasses-based systems [139] allowed people to interact with digital content while on-the-move, without having to use a computer. The systems displayed digital content directly onto the user's glasses, and used a separate one-handed keyboard to interact. However, even Starner concedes that *"potential users avoided [an early glasses-based device] because of perceived social awkwardness."* Others have described how these systems can cause problems with peripheral vision and occupy visual attention [49]. While these heads-up systems do indeed provide always-available information access, it is clear that, at the same time, they remove both focus and attention from the user's surroundings and the physical world around them.

Similar arguments have been made by Jones [72], Turkle [150] and Victor [153], amongst others. In this thesis we consider the loss of immersion in, and engagement with, people and places to be unacceptable unless absolutely necessary. As a result, much of our research is concerned with improving how we interact with the digital content that relates to the physical things that surround us. There is a clear desire by people to interact with this content, but current designs are primarily screen-focused, forcing a comparison between the real scene and the virtual display. While attempts to provide smoother geolocated interaction are becoming more common, existing attempts to merge glasses- or screen-based augmented reality with geolocated digital content break the fluidity of interaction by requiring the user to hold a device as a digital window [85], or wear glasses and a backpack [116].

Lumsden and Brewster [95] have previously motivated moving away from visually augmenting the physical world to mobile devices that are less screen focused. They argue for a shift in mobile interaction to approaches that are less inspired by typical desktop paradigms and aimed more at being eyes-free or hands-free instead. Brewster et al. [15] made a similar early proposal for eyes-free devices, focusing on using 2D and 3D audio to sonically enhance mobile interaction. Their primary motivation was the lack of screen space on mobile devices where input is difficult due to size. While other eyes-free input techniques (such as chording keyboards, for example) were available at the time, Brewster et al. argue that these options require specific learning and can be hard to use. Instead, the authors used a mobile headset, and, to navigate through the spatial audio, users nodded in the direction of the audio they heard.

Using audio or speech for eyes-off interaction can be problematic, however. The main issues, even when discounting recognition problems, are the loss or impairment of normal hearing (due to the requirement to wear headphones), and the social acceptance issues of using speech recognition or publicly audible sounds [15]. Gesture- or touch-based wearable computers have been created in an attempt to address this issue by using, for example, electromyography [31] or wristbands [11, 34], but these approaches require the user to wear custom hardware that is currently unrealistic for everyday usage. Similar approaches are described in less obtrusive belt-based haptic systems – the FeelSpace project [103], for example, aimed to give users direction perception via a vibrotactile belt which constantly indicated the direction of north. Van Erp et al. [44] applied the same concept to navigation, but although the feedback was seen to be appropriate in situations where visual-primary interaction was not suitable (their example focused on military applications), these devices were not appropriate for spontaneous usage unless the belt was worn all the time.

While belt- or wrist-based haptic systems are less intrusive, and more socially acceptable than audio or more obtrusive wearable devices, the requirement to wear extra, single-purpose hardware means that these types of systems seems unlikely to be widely taken up except in specific domains (e.g., [70]), or in cases that offer huge benefits to the target users (such as flight suits, for example). However, recent improvements in mobile hardware and inertial sensors now allow for devices that do not require extra hardware to be worn; instead, the mobile itself can handle all requirements, supporting interaction that is controlled by moving the device itself.

For input, as Oakley and Park [105] have shown, this type of interaction is defined by kinaesthetically-identifiable movements. That is, input can be performed via the user's fluid, implicit awareness of what particular body parts are doing. On mobile devices, this awareness is often achieved by using inertial sensors to detect motion and orientation. For output, while there are many far more unusual and uncommon interactions, such as olfaction [96], electromyography [31], shape changes [59] or weight shifting [58], in this chapter we focus on the types of feedback used in our prototypes.

As we saw in Chapter 1, the typical arguments for heads-up or eyes-free interaction focus around a loss of situational awareness, or a desire to shift users' attention away from screen-primary devices. Li [87] summarises the approach of many current haptic designs: *"haptic research typically focuses on increasing the bandwidth of the tactile communication channel."* That is, rather than focusing on applying tactile feedback to situations where it is most appropriate, quite often, systems are designed to expand haptic feedback to fit *all* situations.

More appropriate applications, in line with our aims in this thesis, can be found in the work of, for example, Strachan et al. [144], who used inertial sensors to recognise when the user positioned a mobile device at specific parts of the body. Their BodySpace system aimed to maintain "natural fluidity" of interaction, allowing users to control a music player by moving the device to, say, the waist or ear, and then tilting the device. Hoggan et al. [61] investigated the use of both audio and haptics, quantifying the limits of these methods, and showing how they can be adopted appropriately in various scenarios. Brown et al. [18] used 'tactons' to define multidimensional tactile icons for prompts when visual interfaces were not available, finding high recognition rates for up to three dimensions of information.

These previous systems give an indication of how non-screen modalities can be used for eyes-off mobile interaction. In the next section we consider how these types of systems, and the interactions they support, relate to each of the contributing chapters in this thesis.

2.3 Discovering and displaying: physically grounded interaction

Many of the early proposals made by Egenhofer [41] for pointing at and discovering geolocated information have since been implemented in mobile device form. Rukzio et al. [123] created several demonstrations of these techniques, and studied touching, pointing and scanning for locating smart objects. They found touching and pointing to be the preferred interaction techniques if the user had a line of sight to or was close to the target device. Pointing was seen as a quick technique that required some cognitive effort but a low amount of physical effort, especially when objects were not within touching distance. Results from their study also showed both pointing and touching to be intuitive techniques, particularly among older participants, who wanted to be able to avoid the need for direct mobile device input as much as possible.

Wasinger et al. [156] created an early pointing-based location interaction system, combining GPS and compass data with speech recognition to allow a user to say a query (e.g., *"what is that?"*) whilst facing a location. Their system processed the speech data, recognising the information request, but did not actually present the user with the requested information. This type of deictic gesture, originally used by Bolt [9] in conjunction with large displays and speech input, is perhaps the most common pointing interaction used in previous work.

Fröhlich et al. [48] conducted a Wizard-of-Oz style user study to assess the viability of these point-to-select interactions against several other methods, concluding that pointing gestures were "highly attractive and efficient" forms of location selection. Building upon this work, Simon et al. [133] described the spatially-aware mobile phone, a conceptual device to connect the physical and digital worlds. Their framework used a three-dimensional model of a location in conjunction with knowledge of a user's position in order to create a line-of-sight visualisation from the user's position.

Continuing this concept, Simon et al. [135] created a point-to-discover application using their earlier framework, using 3D models and position knowledge for a line of sight pointing concept. Their application prototype used location and heading information to, at the push of a button, calculate the visible points from the user's location and display relevant information about them. A further paper by Simon and Fröhlich [132] discusses a similar concept that presents the user with Wikipedia articles about locations near to them based upon their location and the direction they are facing.

Each of these approaches demonstrates active, focused mobile spatial interaction, with the user conjuring up data by actively pointing the device and pulling in content. In contrast, the RelateGateways project [52] used less complex spatial contextual information to push directional information about pervasive services available to the user, including the heading and distance of these objects.

This previous research demonstrates that point-to-select is a viable method of interaction, and can provide users with valuable location-specific information. But while this work provides valuable insights into possible methods and uses of location-based interaction, each also requires virtual location models in order to be able to pinpoint the user's targets. In many of our prototypes we use a similar point-to-select interaction, but in our designs we provide users with similar data without the need for location models and visibility calculations. We imagine situations where such models are unlikely to be created: landscapes, seascapes and very rural areas, for instance, are unlikely to be mapped in great detail in the near future. In addition, our work could allow users to mark objects that are evolving and may never be modelled—live events, shows or funfairs for example—by allowing them to mark any item in the space around their position.

2.3.1 Non-screen-based location-aware interaction

Moving away from visual feedback, Strachan et al. [145] used location and heading data in conjunction with real-time trajectory prediction to guide a user along a path to a desired target location. By pointing and tilting a device around their environment, the user can browse the route features around them, with both audio and haptic feedback directing them toward their destination. When the user is heading toward the target the audio signal is clear and there is no haptic feedback, but if they move off track the audio is distorted and vibrotactile feedback increases. In a related paper, Strachan and Murray-Smith [143] studied mobile interaction with virtual targets, with both vibrotactile and audio feedback. Their research addressed specifically the problems that can arise due to uncertainty in the user's location and heading data, and offered a probabilistic approach to this problem. An experiment using their system showed that targets could be selected effectively, even when fairly tightly spaced.

These systems use gestural point and tilt data from a mobile device in-hand to determine a line of sight from the user's current position. Our design in Chapter 3, however, uses pointing (heading) data to determine the direction the user is facing, and tilting to allow the user to specify the distance of an information point. Similarly, Strachan et al. [144] used location and heading data in conjunction with real-time trajectory prediction to guide a user along a path to a desired target location. By pointing and tilting a device around the environment, the user can browse the features around them, with both audio and haptic feedback directing them toward the specified target. Their system intentionally presents the uncertainty in the system to the user, and allows them to probe possible future routes in the available space, sensing the feedback from routes up to 20 m ahead of their current location.

Holland et al. [62] created *AudioGPS*, using audio to provide representations of the direction and distance of waypoints for navigation. Their backpack-based system was designed to encode direction via panning of a non-speech audio source, and used the speed of audio pulses to represent distance. Their design did not use a magnetometer for distance; instead the system used recent GPS readings to infer the direction of motion. Holland et al. found the system adequate for navigation tasks, and able to offer at least eight separate directions for navigation (i.e., 45° apart). However, due to the lack of an electronic compass the system was unable to offer directions within the first 10 s to 15 s after a direction change.

Jones et al. [76] designed a similar system to help users navigate through a virtual environment using ambient spatial audio, providing cues as to the direction and distance of a specified target by adjusting the fade and balance of an audio track. Brewster et al. [16] used structured audio messages to help users navigate through four levels of an hierarchical menu, finding over 80 % accuracy for location identification. Similar to our haptic discovery interface in Chapter 4, cues helped users discover any currently available navigation possibilities and determine their present location in the menu hierarchy. Our work uses tactile feedback, however, rather than audio icons, to allow users to navigate through a two-level hierarchy of geolocated content.

2.3.2 Delayed interaction

As we have seen, the use of non-screen-primary modalities on sensor-equipped mobiles has previously been used to provide capable interaction methods without the need to focus attention on a screen. One interesting possibility for further extending this is to consider whether there is a need for any screen-based feedback at all.

In Chapter 3 our work is not focused on the real-time delivery of location information, but rather on allowing people to discover situated digital content that might be browsed through at a later time. This delayed approach to information retrieval resembles the concept of "slow technology" (for example, in [53]). Jones et al. [73] present another approach to delayed search in which textual notes jotted onto a handheld device were later used to provide packaged web information via a search engine. More recently, major commercial search engines have provided means for users to read and reflect on their own search histories.[3]

---

[3] Google Web History, for example; see: https://www.google.com/history/

In a development more similar to the prototypes considered throughout this thesis, Rekimoto et al. [118] created LifeTag, a location recording tool using WiFi to find and store a person's position as they travel around a city. As the user moves, their location is tracked automatically, but it is also possible for them to manually 'bookmark' a notable location if desired. Later, when analysing the resulting data, the authors are able to create sculpted views of the visited areas, highlighting in more detail those places where lots of visits occur and shrinking those which are less popular. Currently developed applications for their system include an automatic photo geotagging tool, and playback of a user's journey augmented with related geo-tagged pictures, sourced from Flickr using the locations recorded along the route.

The studies discussed in Chapter 3 look at how content might be discovered, using both visual and non-visual approaches, but do not consider the browsing of this content. In our previous work (see [120]) we have investigated desktop-focused approaches to delayed content browsing, and in Chapter 4 we consider how this content could be displayed and browsed or marked in situ on a mobile device.

2.3.3 Displaying people, rather than media

In the second half of Chapter 4 we turn to investigate how eyes-off feedback might be used to display dynamic content, in the form of a mutually-convenient group meeting point. Similar work has previously used a mobile device to display non-visual instructions to help people navigate. Sokoler et al. [137] describe a low-resolution tactile approach where one of four pegs is raised to give a direction cue. Lin et al. [89] used structured tactons to provide rhythmic vibrations as navigation aids (turn left, right or stop). Their prototype achieved high levels of tacton recognition, finding that users were able to pay attention to their environment at the same time as using the system. However, a Wizard-of-Oz approach was used to guide users to the target location. Our system uses simpler—but realtime—navigational feedback, with the user scanning their mobile device to display the direction they should head in.

Previous work has investigated user behaviours when rendezvousing, but these have used screen-based systems. Axup et al. [5] examined an early prototype which allowed users to send group text messages to co-ordinate a rendezvous. They found that while users were able to meet up, the method used was "somewhat unpopular" and had many usability problems. In addition, and particularly related to our design, they found that the visual attention required to operate the system forced users to alternate their attention between screen and environment to avoid walking into obstacles. Our design uses a more lightweight, low-attention interface that allows users to concentrate their attention on their surroundings rather than a mobile device.

Olofsson et al. [106] studied user needs for meeting during music festivals, and proposed a concept device to display the location of nearby friends overlaid on a map of the festival, based on their findings from a field study. Nicolai et al. [104] explored social contexts in location-aware systems, using mobile proximity-awareness of around 10 m to give a group of users feedback when members joined or left. The system provided no navigational assistance; instead it gave users a brief overview of the people nearby, and classified them into familiarity groups. Our system focuses instead on groups of individuals who want to meet each other over larger areas, guiding them to a convenient meeting point without the need for visual feedback.

Other authors have looked at how people behave during rendezvous. Colbert [29, 30] extensively explored a diary study of users' behaviour while rendezvousing, looking at the effect of factors such as group size, time pressure and area familiarity. Larger group sizes were found to cause more stress to participants, but they were still able to rendezvous successfully. Area familiarity also affected users' rendezvous behaviour: rendezvous in unfamiliar locations required more communication and caused problems attributed to the lack of local knowledge. This aspect is directly addressed by our system – with our design no group communication is necessary, and although local knowledge may shorten the time taken to meet up, it is not required.

Dearman et al. [35] conducted an exploratory Wizard-of-Oz field study to investigate mobile location-aware rendezvous behaviours between pairs of participants. Those using a visual location-aware handheld device chose a meeting location that was a middle point in the majority of cases, with only one pair choosing a landmark. Further work by Dearman et al. [36] investigated user requirements during rendezvous, finding that participants often maintained continual awareness of partner and meeting locations. Our system allows for this by providing meeting point awareness on demand, but does not provide partner locations in order to preserve their privacy during the meetup process.

2.4 Shaping: dynamic location-based interaction while moving

The previous research that we have reviewed so far has generally focused on using multimodal interaction methods to allow removal of the screen element of mobile interaction. In Chapter 5 we turn to investigate how the locations of the users of these devices themselves can be used as part of the interaction. We look more closely at how people's navigational movements and behaviours can be shaped by their digital interactions with the physical world around them.

When navigating using a mobile device, current systems are focused almost entirely on getting the user to their destination via the most direct route. Previous research has investigated the problems with current mobile pedestrian navigation systems, showing how these systems can remove some of the spontaneity of exploration and have an impact on users' normal behaviour [19]. Furthermore, rather than helping people develop an awareness of their location and routes, users of pedestrian navigation systems are often *completely* lost [19]. Seager [127] discusses many of the challenges in screen-based pedestrian navigation, showing how map-based applications are far from ideal.

Holland et al. [62] describe potential problems in navigation, and offer a solution in the form of audio cues to guide users towards a destination. Similar approaches were taken by Jones et al. [76], Strachan et al. [141] and Williamson et al. [159] by dynamically adapting the music that a user is listening to in order to guide them in a certain direction. While these approaches have shown promise, related early work has found that many users are reluctant to use headphones for this type of task [10], citing concerns about being recognised as tourists, or a feeling of isolation from the environment. Our approach in Chapter 5 is to minimise these effects, using vibrotactile rather than visual feedback to allow a less-restrictive interaction style that does not require users to follow turn-by-turn directions.

Previous alternatives to turn-by-turn navigation include landmark-based methods such as that described by Goodman et al. [50], who found benefits in using images of recognisable views along a route to guide users. Krüger et al. [83] and Aslan et al. [4] discuss how users learn routes while using mobile devices, finding that turn-by-turn systems often fail to convey appreciation of the navigation environment to their users.

Our designs take a minimal approach to pedestrian navigation, removing turn-by-turn instructions to prompt users to explore, rather than hurry through their surroundings. We use tactile feedback as a directional cue rather than an instruction, aiming to indicate the ultimate direction that should be taken, but not restricting users' routes.

Previous research has investigated the use of directional vibrotactile feedback as a navigational guide, with vest- or belt-based systems being the most common approach. Van Erp et al. [44], for example, studied several combinations of vibrational pulses, and were able to successfully guide users to walk between waypoints using distance-coded feedback. Further investigation by van Erp [43] showed that well-placed tactors (small vibration motors) could be accurate as directional indicators.

Similarly, Lindeman et al. [90] used tactors placed around the user's torso to help them in a building search task, finding that using directional vibrotactile feedback helped improve users' performance and significantly reduced the number of undiscovered areas. Shin and Lim [131] used a vibrotactile jacket in conjunction with an ultrasound sensor array to provide obstacle detection and feedback for visually impaired users, and they suggest that using haptic feedback for this task can help users to accurately navigate around obstacles without losing track of their path. Luk et al. [94] describe a prototype for mobile haptic interaction where piezoelectric actuators are used to provide several tactile sensations, ranging from simple buzzes to complex patterns.

Pielot et al. [112] used a haptic belt with directional vibration to help users of paper maps orient the map correctly as they walked. The belt vibrated continuously, and users were able to incorporate this background cue into their navigation behaviour. A similar approach was taken by Johnson and Higgins [70], applying the technique to navigation for blind users. Their tactor belt was aimed at helping people avoid obstacles in their surroundings, motivated in part by a desire to lessen the effect of navigation on users' other activities. Our systems in Chapter 5 have a similar goal: allowing interaction with a navigation device to be thought of as a background task undertaken only when it is necessary or desirable, rather than providing feedback for slight path deviations or upcoming waypoints. We aim to promote opportunistic navigation interactions, allowing the user to make casual, infrequent requests for feedback when it is appropriate, rather than being guided constantly toward a target destination.

Lin et al. [89] used a related approach, providing navigational assistance via direction-specific tactons, and finding that users were quite able to recognise the haptic cues and take the indicated paths through their environment. Strachan and Murray-Smith [143] investigated the use of simple directional vibration similar to that used in our systems, applying this to help with spatial target selection. Haptic target finding was studied further by Ahmaniemi et al. [1], using more complex vibrotactile patterns than Strachan and Murray-Smith, though no significant improvements were found when additional cues (such as 'close to target') were added.

Our work in Chapter 5 builds upon these findings, and relates closely to the bearing-based feedback used by Nagel et al. [103]. We combine this with the low-attention feedback aims of Sokoler et al. [137], also drawing upon our findings in Chapters 3 and 4 showing the benefits of handheld directional vibrotactile feedback while moving. In doing so, we are able to provide a new perspective on the task of pedestrian navigation, and demonstrate the potential of these more flexible approaches.

2.4.1 Socially-influenced navigation

In the second half of Chapter 5 we investigate dynamic directional feedback that is a function of social location models. Dourish and Chalmers [40] define social navigation as *"an artefact of the activity of another or a group of others,"* concisely describing this extensive area of research. While much research has concentrated on using social media to navigate within digital data [100], others have explored social aspects of physical navigation and exploration through approaches such as geolocated images [93], collocated users [142], or specific location-based applications (GeoNotes [45], for example). More recently, these social navigation approaches have also started to appear in consumer-level devices. In-car GPS navigation tools now commonly incorporate live updates to allow both official traffic news and feedback from other drivers in the nearby area.

Karimi et al. [79] define a framework for social navigation networks, using a friend-based system to create recommendations for places to go and routes to take. This framework was tested by Kasemsuppakorn and Karimi [81], allowing a group of users to annotate locations and create personalised recommendations for routes and destinations. In a user study, participants were positive about the utility of the system, but the authors concede that further pedestrian path generation methods are needed. In our social navigation design we take a different approach to these personalised recommender systems – rather than filtering social data within a small group of friends, we adopt a similar technique to the serendipitous search query awareness of Jones et al. [74]. Like the mobile query awareness provided by Church et al. [28], our social navigation design provides location awareness from public social media, using this as an ad-hoc replacement for map data, rather than creating specific route or destination recommendations.

2.5 Challenges and considerations

There are many technological, methodological and ethical challenges that have had to be addressed in the course of this work. Foremost for the majority of our designs is the need to enable eyes-free interaction on 'normal' mobile devices. At the time our prototypes were created, standard mobile devices did not have the range of inertial sensors, or the sensor sensitivity necessary for our designs. However, we were careful to use alternative hardware that seemed likely to become available as a standard component of mobile devices.

Consequently, for our prototypes we primarily used a Bluetooth-connected external sensor pack. We chose the SHAKE (Sensing Hardware Accessory for Kinaesthetic Expression, SK6) for its combination of small size, fast response and simple data retrieval API that is available on many platforms. The SHAKE provides realtime three-axis accelerometer, magnetometer and angular rate data. It also contains a pager motor with variable speed control and active braking, which we used to produce the vibrotactile effects in the majority of our prototypes. Further details, and open-source device drivers are available online (see [64]).

2.5.1 Ethical and methodological considerations

The user studies that we explore throughout this thesis have involved human participants, recruited primarily from Swansea University staff and students, and, in Chapter 6 from users of an existing telephone-based service in Gujarat, India. In each case, the associated ethical issues have been carefully considered, and all participants have been made aware of their right to terminate a session or withdraw from a study at any time. Participants completed a consent form at the start of each study session, and all were given a detailed explanation of the study aims, methodology, data collection, risks and their rights during the study. At the end of each session, participants were given a gift voucher as a token of our appreciation.

All of the studies associated with this thesis were evaluated and approved by the Ethics and Risk Assessment Committee in the Computer Science Department of Swansea University.

Walking speed as a measure of usability

For several studies in this thesis we have studied users while moving, which has raised extra challenges. Previous work has shown how walking speed can be a useful proxy measure for the usability of a system. If someone is able to walk at their comfortable natural speed while interacting with a system, this suggests that the interaction is not seriously disturbing their normal moving behaviour. For many of our systems this is particularly relevant, as our aim is to allow eyes-off usage without interfering in people's normal actions.

Several researchers have shown that using a device while moving can have a detrimental effect on user performance and mobility. Mustonen et al. [102] studied users reading text on a device screen while walking, finding that participants' performance decreased while moving. Similarly, Barnard et al. [6] compared user performance when sitting and while walking, finding that moving significantly increased the time taken for word search and reading comprehension tasks. Pascoe et al. [108] looked at usage of a mobile device for fieldworkers, and suggested that the introduction of minimum attention user interfaces and the addition of context awareness could help to improve the effectiveness of mobile interaction in this environment. Oulasvirta et al. [107] investigated attention to the screen of a mobile device in several different environments, ranging from a laboratory to a busy street, finding large differences in user mobility between these situations.

The percentage of preferred walking speed (PPWS) measure has previously been used as an evaluative measure to assess mobile interactions, and Petrie et al. [111] argue that it can be used as a measure of a device's effectiveness. Pirhonen et al. [113] found PPWS to be a sensitive measure of the usability of a mobile MP3 player, where an audio-touchscreen interface affected walking speed significantly less than the standard graphical version. Kane et al. [78] subsequently adapted the technique and used it to test 'walking user interfaces.'

In several of our studies, the traditional methods for gathering PPWS data would have raised issues with accuracy. For example, in Chapter 3, pedometer-based measurement of PPWS was found to be unreliable and inconsistent for shorter distances. To solve this accuracy problem, we placed a tape measure at the side of the circuit that participants walked around, and referred to this at the point of completion.

PPWS has traditionally been used as a summary statistic for a whole trial, but in an outdoor environment, like those in our studies in Chapters 4 and 5, using an aggregate measure would have caused large variance issues. We compensated for the increased variation in the outdoor environment by using a much higher resolution analysis than is commonly used, down to the level of individual steps. These methods, based on those described by Crossan et al. [33], allowed comparisons of behaviour at each different stage and condition in the experiment.

2.6 Conclusions

In this section we have reviewed previous related work and situated this thesis amongst similar research. Our work here focuses around a number of specific themes, and gaps in previous knowledge. Most importantly, we have seen how many of the previous designs for heads-up or eyes-free interaction focus around completely removing the need for screen-primary interaction. Here, and in all of our designs, our aim is for eyes-off interaction only when appropriate, rather than completely eyes-free at all times.

Previous in situ information discovery designs have focused on using three-dimensional model-based target selection for both people and places. Our work in our first two contributing chapters looks at how these interactions can be provided without the need for box or object models of every environment that might be encountered. Furthermore, we extend the reach of previous non-screen designs to demonstrate and evaluate multi-level browsing of digital content hotspots. Previous work has used non-screen methods for the discovery step only; we extend this and show fully haptic-supported interaction and filtering of located content.

In this chapter we have also seen the gaps and shortcomings in current approaches to pedestrian navigation, and how a focus on shortest time or constant location awareness can have the opposite effect. In our work we allow eyes-off vibrotactile discovery of the possibilities for exploration in a user's surrounding area, and also let them display, dynamically, an infrastructure-free, privacy-preserving group meeting point. We offer a new perspective on pedestrian navigation, lessening the need for waypoints or turn-by-turn instructions, and instead aiming to support autonomy and location awareness, with underlying confidence about reaching the goal.

CHAPTER THREE

Discovering

In this chapter we introduce the first of three elements of eyes-off physically-grounded mobile interaction. We investigate whether screen-primary feedback is necessary to support the discovery of geotagged digital content, and subsequently show how inertial sensors and vibrotactile feedback can be used for discovery without screen-focused interactions. This is explored using novel mobile device prototypes, demonstrating empirically the extent to which location-based information can be discovered using approaches that are less focused on screens than many current research or commercial methods. In doing so, we can begin to outline important properties for designs of similar devices in the future.

3.1 Introduction

There is an increasingly rich and large set of geotagged information available while mobile. For example, people regularly share location-tagged status updates and photos using social networks; there are also encyclopaedia articles, current event schedules, music, videos, maps and so on. Some of this information is explicitly about the place – for example, a description of a nearby historic building, or a picture of a scenic beach. There is also further information that can be automatically associated with a location from the behaviours of people as they pass through, such as the music listened to by others in an area, or previous searches by other visitors (e.g., [74, 75]). Increasingly this content is being discovered, collated or created entirely using mobile devices, rather than laptop or desktop computers.

As we discussed in Chapters 1 and 2, while current visually-oriented views of this content work well on the large fixed screens of remote, conventional computers, when these methods are used in situ on mobile devices, a focus on the screen can lead to a loss of touch with the physical places that the content is linked to [72, 83]. We argue that innovations are needed for effective, more personal mobile discovery of this digital content. In this chapter we are motivated by a desire to provide people with eyes-off, non-screen-based ways of discovering digital content associated with the places they live in, visit, or perhaps just pass through occasionally.

Our approach is to use the mobile phone as a pointing device to discover, select or save geotagged content associated with physical objects or locations. Point-to-select gestures such as these have previously been employed to let people use the affordances of the physical world to directly access the digital content around them (e.g., [48, 132]), but these have focused on the visual aspects of discovery. Here we investigate point-to-select gestures as a method for providing eyes-off physically-grounded interaction, via two separate mobile prototypes.

Before investigating entirely eyes-off information discovery, it is important to first establish the effectiveness and boundaries of current screen-based methods. Previous work has looked at the use of visual (e.g., [132]) or completely non-visual (e.g., [134]) feedback when pointing to buildings or other objects. However, previous comparisons between these have used Wizard-of-Oz approaches [48] or focused on carefully 3D-modelled environments [134]. Previous work has not determined the extent to which visual feedback is necessary for allowing effective discovery of geotagged information via pointing.

Consequently, our first prototype investigates several different screen-based pointing methods to determine the extent to which screen-primary interaction is required for effective discovery. Its three modes give the user progressively sparser elements of on-screen feedback, and its evaluation helps us to develop an understanding of the accuracy and efficacy that is achievable in these situations. The results indicate that while on-screen feedback is more precise than eyes-off methods (as expected), approaches that do not focus on a screen seem to show promise if their accuracy can be improved.

Our second prototype builds upon findings from the first, removing visuals entirely and instead aiming to improve accuracy by augmenting pointing-based object selection with vibrotactile feedback. This contribution is our first step in moving towards allowing effective eyes-off interaction. We evaluate this second prototype system's efficiency and efficacy in a user study that looks specifically at its usage while participants are walking – an important consideration to be aware of when designing this class of device.[1]

Together with our first study, this experiment helps to gauge the reactions of users to these new interaction methods, and to measure the success of our designs for supporting eyes-off in situ information discovery. From these results we are able to outline several design recommendations for the discovering stage of eyes-off physically-grounded interaction.

3.1.1 Pointing for discovering

The following example scenario illustrates how a pedestrian—Alex—might use his mobile device, eyes-off, for location-aware digital information discovery while on-the-move.

> *Walking around town, Alex holds his mobile phone in hand. The street is full of life and he's enjoying people-watching and checking out the shop window displays. With his hand at waist level, he loosely points toward a bookshop in the distance; he holds the phone almost horizontally as the shop is quite far away. There's no response from his device, so he sweeps it over to his right, to a music store over the road. He holds the phone almost vertically as the shop is so close by: the device vibrates, telling him that there's new music from artists he might like. Alex doesn't feel like stopping just now, though, so he carries on walking – he knows his mobile has saved the sample tracks, so he'll listen to a few previews later on when he's at home . . .*

This scenario illustrates two key interaction behaviours that are explored in this chapter. Firstly, Alex is pointing to physical places or objects using his mobile device, eyes-off, in order to request or save digital information about them. Secondly, tactile feedback given by his device helps him determine where he can find information, again without the need to look at his screen. Over the rest of this chapter we examine these two elements of this interaction in detail.

---

[1] See Chapter 2, Section 2.5.1 for a discussion of previous related research into mobile device usage while moving.

When designing interaction with geotagged content, the initial, and perhaps most basic, element of the process that must be considered is the selection of the object about which information is desired. Screen-based methods often use simple buttons or icons on the display, but recently, direct pointing to physical objects—point-to-select—has begun to move into the real world. Several authors have discussed point-to-select implementations for retrieving geotagged content, as we reviewed previously in Chapter 2. Particularly relevant to the techniques we investigate in this chapter is the research of Fröhlich et al. [48] and Simon et al. [135].

Fröhlich et al. compared four methods for accessing geo-spatial information – point-to-select gestures; a north-up map; an egocentric (i.e., rotating) map; and, an augmented reality camera view. Pointing was ranked as the best approach by the majority of participants, and was commonly viewed as a very efficient gesture. Simon et al. focused on a pointing gesture for interaction with physical objects. Their Point-to-Discover platform is a server-based database of location model data, which is accessed by a mobile client that uses point-to-select gestures to query. Content is retrieved when a gesture intersects with a modelled building or other object. Both of these implementations differ to our approach, however, in that they use three-dimensional location models to interpret gestures and retrieve relevant content.

With systems that rely on such models, a pointing gesture allows the user to indicate a particular direction from their position, and the model can then be used to offer a selection of geotagged objects that intersect with the gesture in that direction. While many popular and interesting places have begun to be modelled in this way in recent years, here we also want to account for situations where it is unlikely such models will be created. Rural landscapes, seascapes and many places in developing regions, for instance, are unlikely to be mapped in great detail in the near future. But there are also locations that may never be completely modelled, such as very remote areas, and the locations around live events, with their constantly evolving centres of attention.

When no environment models are available, there are two options for using point-to-select as an eyes-off interaction style. The first is to provide the user with a method for specifying the distance of their selection (e.g., [120, 145, 159]). Alternatively, the system can simply select the nearest object in the direction the user points, up to some maximum range (e.g., [133, 134]). Both of these approaches treat the content itself as a 'model' of the location – each item becomes a point that can be interacted with, removing the need for a three-dimensional representation of the environment. However, there is no support for occlusion (by physical, rather than digital, objects), as the system has no way of knowing whether the user can actually see the item they are pointing toward.

Providing a distance selection method allows point-to-select interaction to be used to filter and retrieve content more closely relevant to the user's intended target. However, while this approach appears to allow precision in selecting locations, it requires a more complex initial pointing gesture, as the user must be provided with a way to specify the distance of the object they are attempting to select. If this refining action is too complex, or inaccurate, the user will have problems with the uncertainty caused by the complexity of the distance gesture.

Intersecting the user's pointing gesture with the closest available content in that direction allows potentially simpler interaction for the user, but can create ambiguity and uncertainty when discovering content. For example, if nothing is discovered after pointing to an object, it is difficult for the user to determine whether this is because they have pointed in slightly the wrong direction, or the object is out of range, or because there is no information available about the object they have pointed toward.

In this chapter we investigate both of these point model approaches, via two separate prototypes. In our first prototype we add an additional tilting action to an initial point-to-select gesture to allow users to specify and refine the distance of their discovery query. Our approach to this problem is similar to that of Williamson et al. [159], who used location and heading data coupled with tilting to allow users to explore a real-time trajectory prediction. Our use of tilting is purely to allow users to specify an approximate distance for their selection, however, and is intended to be a quick additional refinement, rather than a continuous interaction. We test several versions of this prototype to explore the need for screen-based feedback.

In our second prototype we replace the tilting gesture with tactile feedback. In this prototype the system selects the closest geotagged object that intersects with the user's selection, providing vibrotactile feedback in an attempt to improve targeting accuracy and remove some of the uncertainty in discovering geotagged content. The tactile feedback provided is similar to that of Strachan et al. [145], who used increasing vibration intensities for navigation to warn users when they moved away from their intended route. Here, however, we use feedback intensity to guide users' pointing gestures toward the centre of their intended target to help them infer which real-world object they are selecting.

3.2 Exploring the need for on-screen feedback

As we have seen so far in this thesis, the current commercial state-of-the-art for in situ information discovery is primarily screen-based. Eyes-free or heads-up approaches in the research literature focus on completely removing visual attention; here we aim for eyes-off interaction methods, using screens only when necessary. But before developing these new methods, it is important to understand the extent to which on-screen feedback is required in these scenarios.

In this section we describe the design and evaluation of a prototype constructed to measure the accuracy of three progressively less-visual object selection methods, in an effort to demonstrate that screen-primary methods are not essential to allow discovery of geotagged content. In doing so, we are able to gauge the potential for place-based interaction focused on the place rather than the screen.

3.2.1 Prototype design

We developed a prototype that uses a pointing gesture to indicate direction, combined with a simultaneous tilting gesture to allow specification of distance. This technique allows discovery without environment models, and in a single movement. Our prototype uses a standard Dell Axim x51v PDA, securely attached to a SHAKE[2] motion sensor pack, which provides inertial sensor data. This allows us to capture realtime movements made by the user while holding the device. While it is to be expected that this method is unable to attain the precision of touch screen or model-based approaches, our hypothesis was that sufficient accuracy could be achieved, even without precise visual feedback.

When using the prototype, the user first points in the direction of the location or object they wish to discover information about. Then, to select places furthest from them the device is held flat; for places closer, the device is tilted back towards

---

[2] See Chapter 2, Section 2.5 for more details.

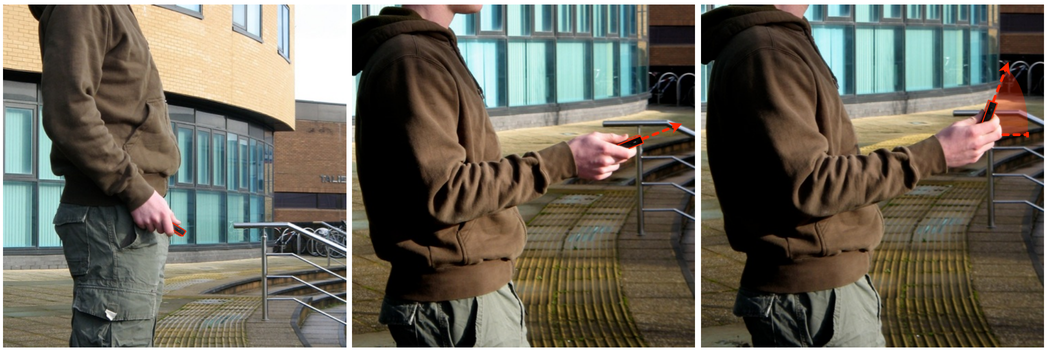


Figure 3.1: Specifying the distance of a selection. Left to right: the user takes the device (highlighted in each image) from their side, points in the direction of their selection and then tilts back to approximate the target's distance. The angle of tilt (shaded in the rightmost image) specifies the distance being selected.

their body. That is, when held horizontally (at 0° relative to the horizontal plane), the distance selected is the maximum range; when it is held vertically (at 90° relative to the horizontal plane) the distance is set to 0 m. Between these two extremes the distance is a linear function of the degree of tilt. A separate Bluetooth GPS receiver provides realtime location data, and this is used in conjunction with the directional and tilt gestures to calculate the target location that the user is pointing to. Figure 3.1 illustrates this interaction method.

Varying on-screen feedback

In order to measure the need for, and the effect of, screen-based feedback, we developed three different interaction modes for the prototype system. Each of the three modes uses the same point and tilt interaction method for selecting locations, but presents a different on-screen display to the user during the process (see Fig. 3.2).

The most screen-focused mode, the *aerial mode*, displays an aerial photo of the user's current location, overlaid with an arrow showing the direction in which they are currently pointing the PDA. The arrow pinpoints the location that can be marked, and adjusts in length and direction as the user moves the device. The aerial photo stays in the same orientation regardless of device orientation – only the arrow rotates. This design choice was influenced by previous work, which found that rotating the view itself added confusion and did not help users [2, 128]. To select locations of interest, users point directly at the target with the PDA in their hand, which rotates the arrow to the correct direction on the aerial photo. They then tilt the device toward or away from their bodies to refine the targeting, which makes the arrow shorter

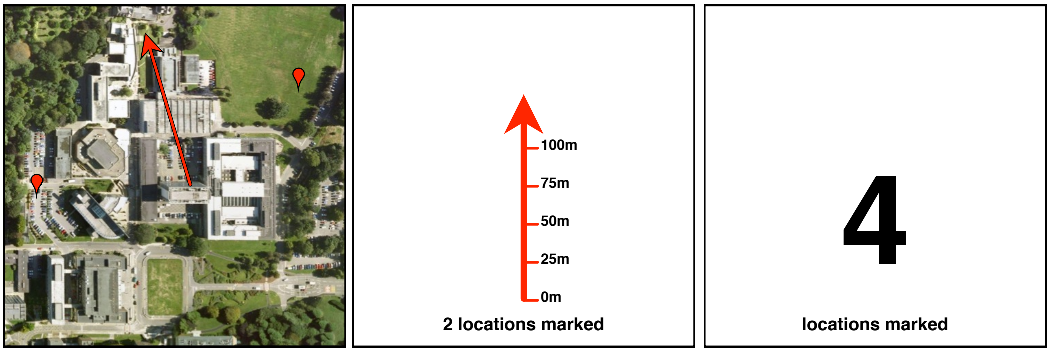


FIGURE 3.2: PDA screen displays from the three system modes. Left: *aerial mode*, an aerial photo overlaid with a moving (rotating) arrow and two locations marked. Centre: *arrow mode*, a simpler arrow-only view with the user currently marking a location approximately 125 m from their present position. Right: *eyes-off mode* – only the number of locations marked so far is displayed.

or longer. When the arrow pointer is positioned over their desired target, the user presses a button to select that place, and an on-screen marker is dropped at that location. This mode, then, provides the user with an accurate view of their selection target, but in a screen-primary manner.

Our second mode, the *arrow mode*, dispenses with the aerial photo display, and instead only shows the user an arrow, the length of which indicates the distance of the location they are targeting. Distance indicators (in metres) are displayed alongside the arrow in 25 m increments. Turning to face a new location does not adjust the rotation of the arrow; instead the arrow always points toward the front of the PDA (i.e., the direction in which the user is pointing). This mode aims to require less visual attention than the aerial mode during selection, and our intention is for users to be able to mark locations without concentrating their attention on the screen except for when refining the distance selection.

The least visually-focused approach, the *eyes-off* mode, removes all indication of distance and heading. Instead, the simple display shows only the number of objects marked so far, and the user must estimate distance entirely without feedback. Our aim in this mode is for the user to be able to select and mark locations without the need to look at the device at all – visual feedback is minimal and provided only for reassurance that a target has been selected and the point-and-tilt gesture recognised.

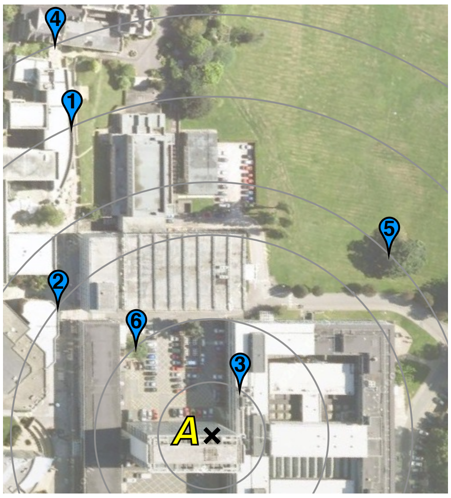
In each of our three test systems, the gesture used to mark target areas of interest is identical, regardless of the display on the screen. Users press a button to start the gesture, and point at the target with the PDA in their hand, tilting the device toward or away from their bodies to refine the distance of their selection. When they feel they have indicated the approximate distance to their desired target location, the user presses a button to mark that place. Each system has a maximum range of 175 m. This upper limit is based on our earlier work (see [120]), which found that the majority of locations marked were within this relatively short range.

The PDA records timestamped accelerometer and magnetometer data continuously while in use, and also marks where the user pressed the button to begin, and then to finish selecting a target. A separate PC-based parsing application is used later to analyse logs from the PDA and extract the coordinates of the places the user marks.

3.2.2 Experiment: exploring the effect of visual feedback on selection

We undertook a lab-based study to evaluate and compare each of the three target selection modes with regards to efficiency and accuracy in marking locations. Tasks and measures were focused on the targeting accuracy that is afforded by each interaction method, allowing us to gauge the extent to which visual feedback is necessary for discovering geotagged objects. We also observed participants' methods of using the prototype, and compared the times taken to mark locations between each mode, in an attempt to determine whether eyes-off interactions could allow easier or faster information discovery. To ensure that the system would work consistently indoors for the study, we chose to remove the GPS element of the prototype, and instead fixed the location of the system to the coordinates of the position where participants would be standing inside our lab.

Thirty-eight participants aged from 18 to 65 were recruited for a 10 min laboratory study. Sixteen participants were university staff members, 22 were students; 15 participants were male and 23 female. Nine participants had previously used sensor-based interaction with a games console; the remainder had no prior experience of this method of interaction. Participants were not asked specifically about sensor-based interaction on their personal mobile phones as, at the time of the study (mid-2007), phones with inbuilt motion sensors were not as common or widely available as they are at the time of writing this thesis. At the time of the study, mobile phone models that were available with motion sensors primarily used this for screen rotation or camera orientation rather than the more advanced interactions (such as game control or camera augmentation) that are now common with current devices.



Target	Description	Dist. (m)	Angle (°)
1	Large event banner on building gallery	140	−22
2	Advertisement flag on side of theatre	81	−58
3	Machinery on roof of adjacent building	23	+27
4	Archway leading to campus exit	174	−19
5	Large tree on main campus lawn	108	+31
6	Large tree in courtyard car park	46	−42

FIGURE 3.3: Left: the six targets used for our study, labelled in the order they were used. Point *A* shows the position of the participant in the lab. Right: distance and viewing angle (from participants' viewpoint, perpendicular to lab) of each of the target points.

Procedure

At the start of the study each participant was met individually and given an introduction to the equipment and its purpose, followed by a demonstration of its use via a short usage scenario example. As a form of training, participants were then asked to familiarise themselves with the system by marking up to three locations of their choosing from a window in our laboratory, until they felt they were comfortable using the system for the study.

After this initial training, participants were then asked to mark six pre-set targets from a fixed location in the lab. These targets were recognisable objects chosen before the study specifically to be at an approximately evenly spread range of distances between the minimum and maximum marking range of the system, and also in several different viewing directions. Figure 3.3 shows the positions and descriptions of each of the six targets, and the order in which they were selected during the study. The points used were all clearly visible from the laboratory window and all participants stood at the same position in the lab during the study. Participants were instructed not to spend excessive time aiming for extreme accuracy; instead they were asked to imagine casually marking a location while wandering around a new place.

Thirteen participants used the aerial mode for training and during the study. A further thirteen participants used the arrow mode for training and the study. The

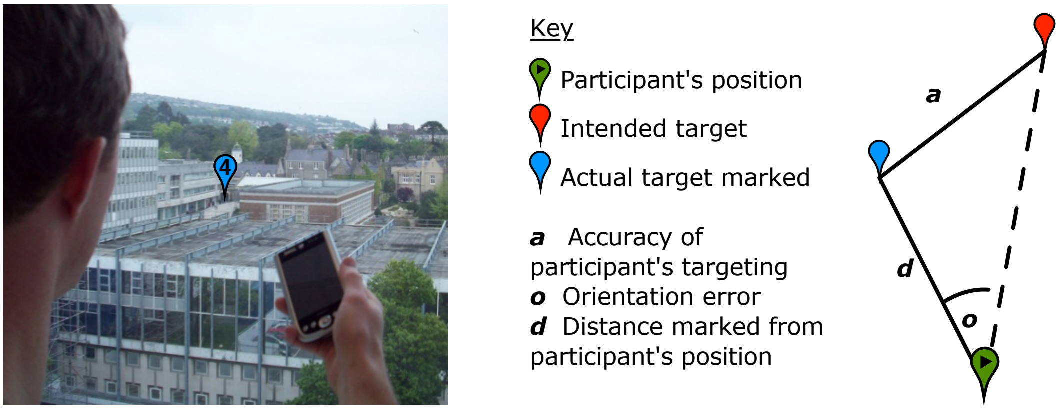


FIGURE 3.4: Left: marking a target from the position participants stood at in the lab. Right: accuracy measurements recorded for each point marked by participants during the study.

remaining twelve participants used the arrow mode for training and the eyes-off mode for the actual study. Our reason for using this method was that the eyes-off mode gives no indication of the distance of the location being marked. We anticipated that had we asked these participants to use the eyes-off mode without any prior knowledge of the tilt required for the distances they would be marking, the results would most likely be hugely variable and ultimately of no use. As the method of marking locations is identical for each mode—only the display is different—we did not expect any adverse effects on our results from this choice of training method.

After participants had marked all six targets, a short semi-structured interview was conducted to gather feedback. Participants were then given a bookstore gift voucher as a token of our appreciation.

Measures

We observed and noted the approach taken by each participant while they used the system, and analysed the system's logs to measure the time taken to mark each target. In addition, three accuracy-related measurements were recorded, as illustrated in Fig. 3.4. These measurements were:

Accuracy The absolute distance (in metres) between the participant's marked point and the intended target, allowing us a general measure of marking accuracy in each mode.

Orientation The angular difference between the actual direction (i.e., compass heading) of the target and the direction the participant pointed in, allowing us to determine the extent to which accuracy errors were a result of inaccurate pointing (i.e., pointing in the wrong direction).

Distance The distance of the locations each participant marked from the point they were standing in the lab, which allowed us to measure both the precision of the tilt estimations required to mark locations, and the effect of this precision on the overall accuracy in marking.

Analysing these components separately allowed us to determine which part of the point and tilt interaction led to inaccurate target selections. For example, a participant may have pointed in the right direction but underestimated the amount of tilt required to mark the location, giving a large distance error, but a small orientation error.

3.2.3 Findings

All participants used the system to mark at least one location during the pre-study training session. During the study, each participant performed the point and tilt gestures to mark each of the six specified points during the study without needing any assistance from the researcher. These marked targets were analysed for each separate mode to allow comparison of the three accuracy measurements and time taken to mark a location between modes, and between each of the targets.

Tables 3.1a to 3.1d show the mean and standard deviation of each measured value over all participants per target and mode. Figure 3.5 shows a visual representation of the approximate total target areas marked over all participants, highlighting clearly the accuracy differences between each of the system's three modes.

When comparing the complete set of marked targets from each mode we can immediately see that participants using the aerial mode appear to have been more accurate in perceiving and selecting targets. However, the time they needed to mark these targets was longer. Participants using the arrow and eyes-off modes took less time to mark a target than those using the aerial mode, but the absolute distance error and angle errors produced were higher. Further statistical analysis of each measurement, in the next section, shows several significant differences between the three modes.

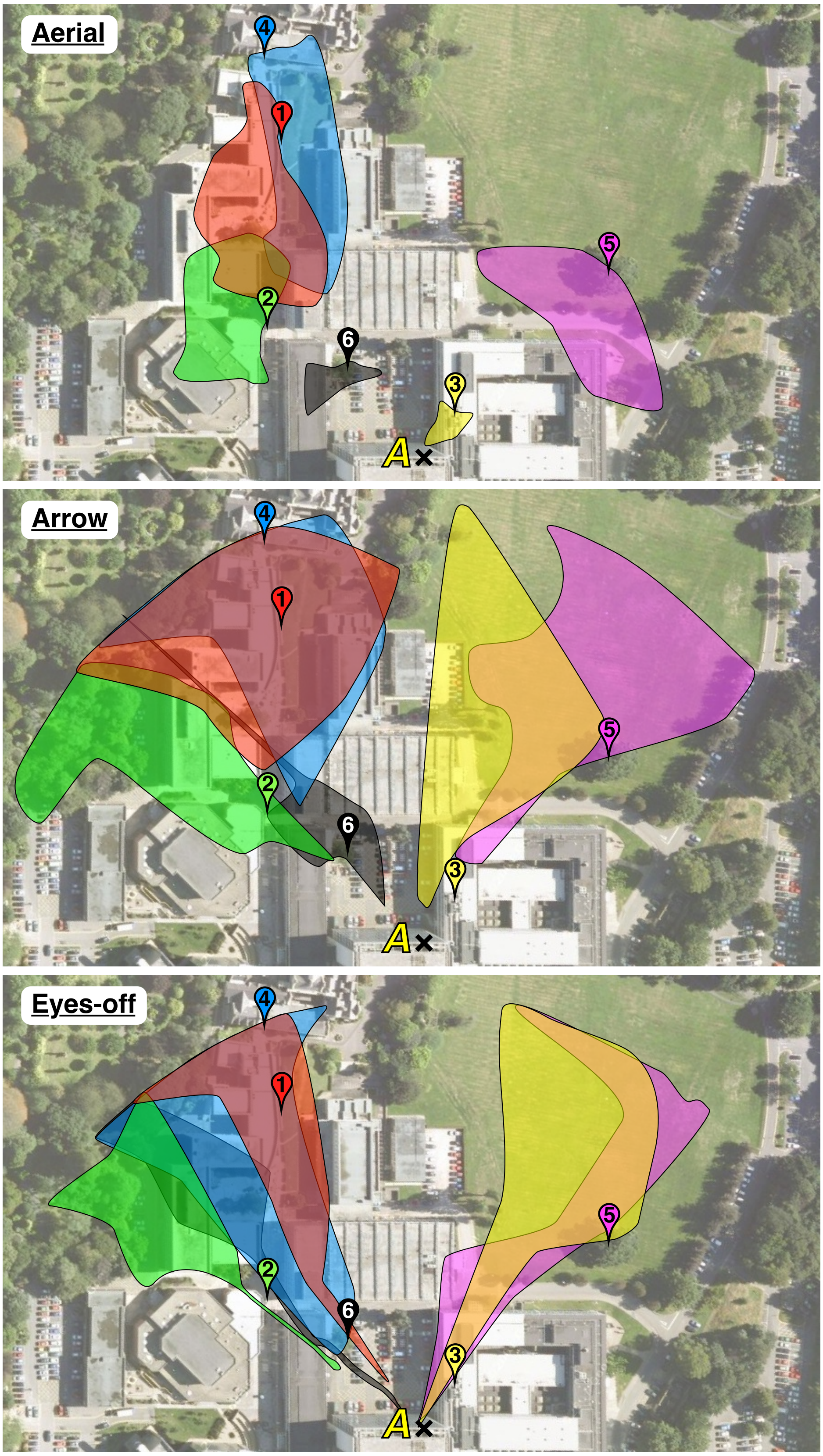


FIGURE 3.5: Targeted area ranges for each mode, showing the intended targets and the outer edge of the actual target areas selected over all participants. Photo orientation is the same as that shown by the system when in the aerial mode. Top: aerial mode; Middle: arrow mode; Bottom: eyes-off mode. Point *A* shows the position of the participant in the lab in each case.

Target	Aerial		Arrow		Eyes-off	
1	38.8	(22.0)	41.0	(20.6)	55.6	(26.7)
2	23.3	(12.9)	55.8	(28.6)	50.0	(29.6)
3	11.9	(4.3)	41.2	(54.5)	96.9	(52.2)
4	42.5	(37.0)	45.0	(29.6)	48.6	(43.9)
5	26.3	(18.2)	62.8	(18.0)	67.3	(30.0)
6	8.6	(6.5)	33.5	(39.0)	69.0	(48.1)
All	25.2	(23.1)	46.6	(34.3)	64.6	(41.7)

(a) *Accuracy* error (m (s.d.)) for each target. On average, the positions marked by participants using the aerial mode of the system were closer to the intended target than those using the arrow or eyes-off modes.

Target	Aerial		Arrow		Eyes-off	
1	14.2	(13.7)	14.6	(10.6)	11.4	(5.9)
2	17.2	(10.5)	14.7	(10.2)	8.8	(6.2)
3	8.1	(4.8)	16.2	(6.9)	31.1	(33.5)
4	7.7	(6.7)	14.8	(10.1)	14.3	(13.2)
5	6.7	(8.0)	13.3	(7.9)	12.2	(9.2)
6	13.8	(11.9)	8.5	(7.3)	9.8	(5.8)
All	11.3	(10.2)	13.7	(9.0)	14.6	(17.0)

(b) *Orientation* error (° (s.d.)) for each target. While some targets were less accurately oriented to by participants than others, no single mode allowed higher pointing accuracy than any other.

T (dst.)	Aerial		Arrow		Eyes-off	
1 (140)	123.3	(26.8)	145.3	(28.7)	122.1	(55.8)
2 (81)	90.2	(16.4)	125.1	(39.5)	107.1	(51.7)
3 (23)	13.8	(7.8)	53.0	(59.4)	107.1	(69.6)
4 (174)	137.1	(37.5)	156.0	(30.6)	148.1	(47.0)
5 (108)	92.0	(10.6)	113.3	(47.4)	111.2	(67.8)
6 (46)	45.6	(4.9)	72.7	(43.5)	100.6	(64.2)
All	83.7	(47.3)	110.9	(55.7)	116.0	(59.9)

(c) *Distance* marked (m (s.d.)) from participants' location, compared to the actual distance of the target. Participants using the arrow and eyes-off modes tended to mark distances further away from where they were standing, regardless of the actual distance of the target.

Target	Aerial		Arrow		Eyes-off	
1	25.9	(11.5)	11.3	(4.1)	9.9	(6.6)
2	27.0	(16.4)	9.5	(3.6)	6.9	(5.5)
3	21.8	(13.2)	7.7	(3.2)	4.6	(3.5)
4	23.9	(14.2)	6.6	(2.5)	5.9	(1.9)
5	20.8	(13.1)	10.2	(5.0)	7.0	(5.7)
6	10.2	(5.9)	7.8	(3.1)	7.7	(9.0)
All	21.6	(13.6)	8.8	(3.9)	7.0	(5.8)

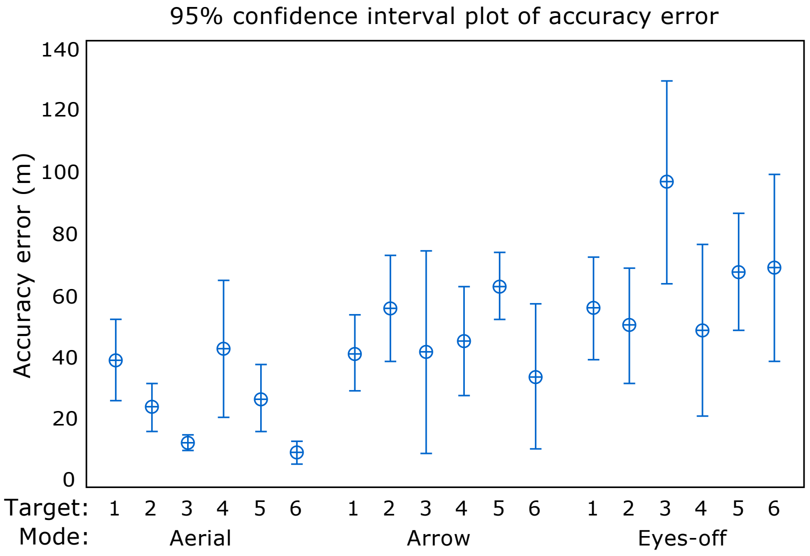
(d) *Time* taken (s (s.d.)) to select each of the targets. Participants using the aerial mode took a longer time to select each target than those using the arrow or eyes-off modes, on average.

Table 3.1: Summary of average results from each of the three accuracy-related measurements, and the time taken to select targets, for all six targets over each of the system's three modes.

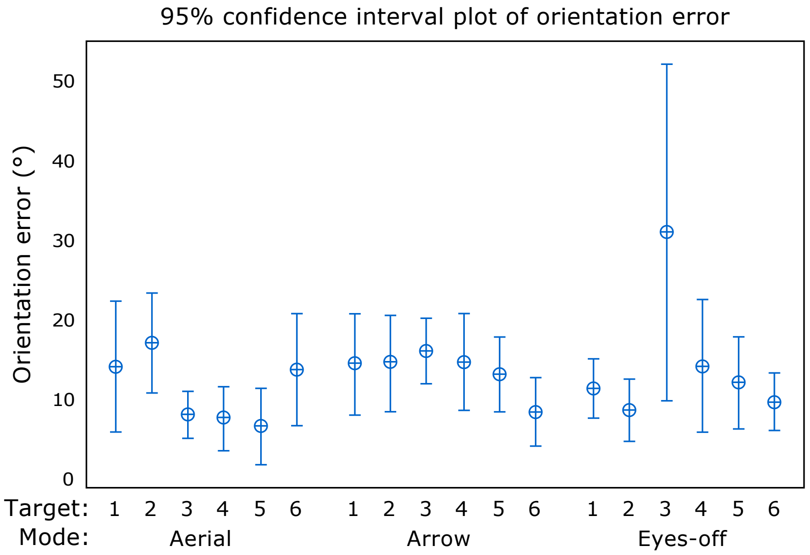
Accuracy

Figure 3.6a illustrates the spread of target selection accuracies recorded. Using GLM ANOVA it was found that the type of feedback used (i.e., the system mode) had a significant effect on the accuracy error ($F = 25.66, p < 0.05$). There were, however, no significant differences between the six individual targets.

Posthoc analysis using a Tukey test shows that there is a significant difference in accuracy when using the aerial mode ($p < 0.042$), indicating that this mode allowed users to be more precise in their overall marking of each target. However, the accuracy errors were not significantly different between the arrow and eyes-off modes ($p > 0.3$), showing that neither of these feedback methods was more accurate than the other. The Tukey test also shows that with the aerial mode, participants were able to mark



(A) Accuracy in selecting each target (m).



(B) Orientation error for each target (°).

Figure 3.6: 95 % confidence interval plots showing the accuracy and orientation measurements for each target in each mode.

the closest targets (3 and 6) more accurately than the others. There was no significant difference between any individual targets when using the arrow or eyes-off modes.

Orientation

Figure 3.6b shows the spread of orientation errors for each target using each mode. ANOVA between targets shows a significant effect on the orientation error ($F = 15.45, p < 0.05$), indicating that some of the targets used were more difficult to point to than others. However, the direction of this error (i.e., to the left or the right of each target) was not significantly different between targets ($p > 0.3$): there is no evidence to suggest that targets to the left or to the right are easier to select. However, we did not record which hand participants used to hold the system, so cannot elaborate on this result.

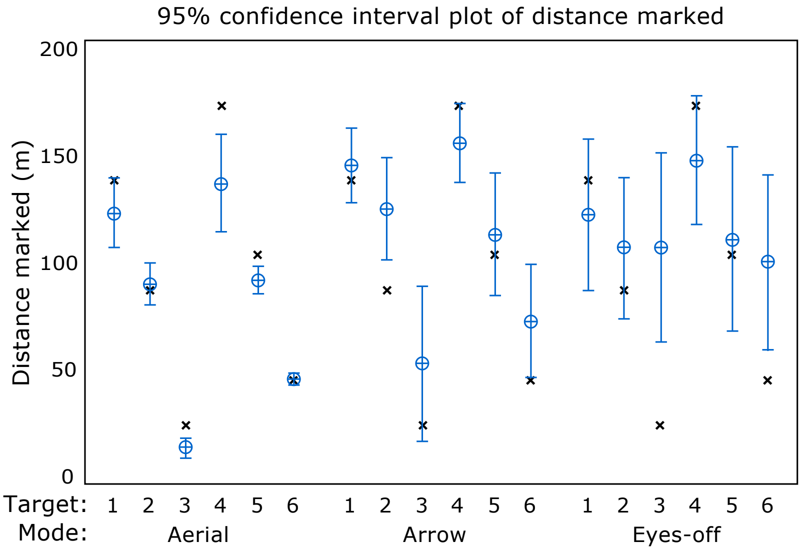
The mode used showed no evidence of a significant effect on either the magnitude of the orientation error, nor the direction of this error to either side of a target ($p > 0.2$): the angular errors recorded for each target were not related to the mode used. In addition, these orientation errors were not significantly different between each system mode ($p > 0.5$), indicating that errors in participants' pointing accuracy were not caused by one particular mode more than the others.

A post-hoc Tukey test shows that the orientation error for target 3 using the eyes-off mode is significantly larger than that for any of the other targets. This error, which is not present for either of the other modes, might be attributable to the complexity of marking a very close target with no feedback for either distance or direction. Target 3 was the least accurately marked position using the eyes-off mode, despite being one of the most accurately marked using the aerial view.

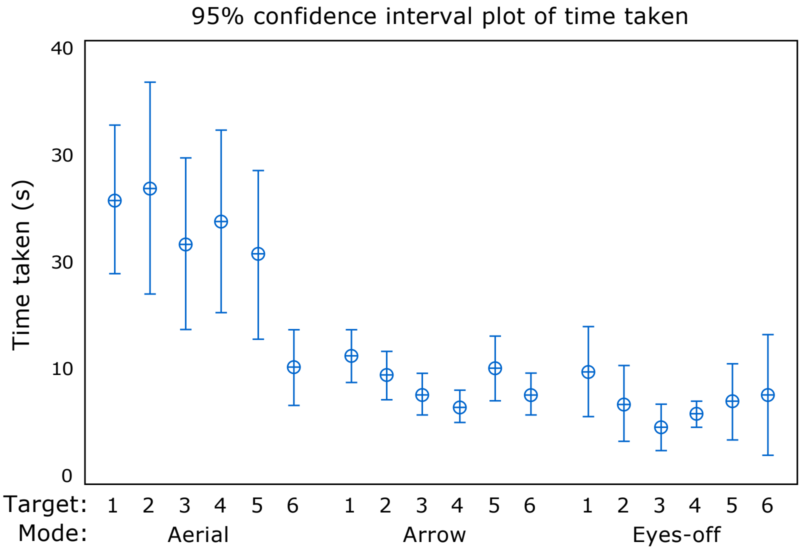
Distance

Figure 3.7a shows the range of marked distances from the participants' position in the lab, with the actual target distances indicated. ANOVA between targets shows that the distance of the target from the participant's location has a significant effect on their accuracy error ($F = 19.04, p < 0.001$): targets further away from the participant were marked less accurately. In addition, the mode used has a significant effect on the distance the participant marked from their position ($F = 7.82, p < 0.001$): participants using the arrow mode and eyes-off modes often marked locations further than was necessary.

A post-hoc Tukey test shows that there is a significant difference between the distance of the points marked from the participant's position when using the aerial mode ($p < 0.02$): the distance marked when using the aerial mode is generally lower than those of the arrow and eyes-off modes, regardless of the actual correct target distance.



(A) Distance marked for each target (m). Actual targets are shown as ×.



(B) Time taken to mark each target (s).

Figure 3.7: 95 % confidence interval plots showing the lab distance and time taken measurements for each target in each mode.

Time

Figure 3.7b illustrates the times taken to mark each target using each mode. ANOVA shows that the mode used has a significant effect on the time taken to complete the task ($F = 27.62, p < 0.001$): the aerial mode is slower than the arrow and eyes-off modes. There were, however, no significant differences in times taken between the six individual targets ($F = 2.062, p < 0.07$): no single target was quicker to mark than any other.

Posthoc analysis using a Tukey test shows a significant difference between the times required to complete each marking task in the aerial and arrow mode conditions ($p < 0.001$), with those participants using the aerial mode taking significantly longer to mark each target. The test shows no evidence of a significant difference between the times taken to mark each of the targets when using the eyes-off mode ($p > 0.35$).

Participant feedback and observations

In the post-session interviews all participants agreed that they found pointing while looking at a location to be a natural way to request information. Recorded observations of participants' approach to the task indicated that participants using the eyes-off mode and, to a lesser extent, the arrow mode, were able to mark locations without needing to look at the screen. Participants using the visual mode generally spent a large amount of time comparing the view on the screen with the view in front of them, resulting in higher accuracy but longer task completion times.

Participants using the eyes-off mode appreciated the low effort required to select a target, but several commented on the lack of accuracy in their marking (even without feedback on their eventual accuracy), finding it difficult to estimate the tilt required to specify distances. Two participants using this mode said that after using the arrow mode for their training they were confident selecting targets at various distance intervals, however three other participants found that they were unable to recall the distances used in their training, so had to guess at the tilt required during the study.

Participants using the arrow mode were positive about the ease of marking locations, and several stated that it was easy to mark targets. However, similarly to the eyes-off mode, a number of participants remarked on the difficulty of estimating distances accurately from their location, primarily due to the complexity of estimating a visible distance from their viewpoint as a figure in metres. Other minor issues were raised by participants using this mode, such as height differences between their location and the target positions, and the difference in viewpoints depending on each participant's eye level.

Participants using the aerial mode were confident in their ability to accurately identify a target on the aerial photo of their location. Most participants subsequently found that precise and accurate marking of each location was more difficult in practice, due to the need to constantly compare the real view in front of them with the aerial photo on the device. Five participants spent a large amount of time (over 40 s per target) making their gestures very specific. The maximum time taken for each of the other modes was 25 s, with the majority of the participants using these modes taking far less time than this.

3.2.4 Discussion

The three progressively less-visual modes used in this prototype have each allowed participants to select targets using pointing and tilting, but the effectiveness attained has varied. The ability of participants to select targets is significantly different between the three modes, as expected. While the aerial mode allowed participants to select targets most accurately, it did so at the expense of time taken to complete the task. The arrow and eyes-off modes were less accurate for selecting targets than the aerial mode, but allowed participants to select targets faster. On-screen feedback clearly helped participants to be more accurate in selecting targets.

When looking at the distance specified by the tilt gesture, all three modes of the system are progressively less accurate for targets further away from the participant's location. When using the aerial mode, targets closer to the user appear to have been easier to mark accurately, a result that is especially clear from Fig. 3.5. One possible explanation for the inaccuracy in selecting targets further away from the users is a lack of understanding of the tilt angles that were necessary to mark a location, perhaps as a result of the small amount of training participants were given. Another possible reason for this error could be a more general lack of distance perception or, rather, an inability to convert visual distance perception into a discrete value in metres. This result is particularly important when considering potential usage of the tilting method as a technique for eyes-off distance selection.

When using the aerial mode, trying to match up the photo with the physical surroundings is a potentially fiddly task, taking the user's attention back and forth between the environment and the screen. Our aerial view's design was influenced by the results of previous work (e.g., [2, 3]), which showed that systems in which a comparison is required between a map and the environment should align maps 'forward-up.' However, as recommended by Aretz [2], we did not continuously rotate the on-screen map to maintain this alignment. This design choice may have been the cause of the additional time taken during the tasks while using the aerial view – participants needed to mentally rotate the map to match the target direction, rather than rotate the device to adjust the view.

The arrow mode's simpler display lessens the need to divide attention between the screen and the view of the surroundings – its partially eyes-off approach shows promise for future approaches that aim to minimise screen-focused feedback. The results from the eyes-off mode are the least accurate of the three options, allowing only gross indications of distance, such as 'near,' 'middle' and 'far' (cf. [24]). Clearly, the lack of feedback has affected participants' ability to mark locations accurately. Despite this inaccuracy, however, the eyes-off system did allow participants to specify direction and some degree of distance without a visual display. If its accuracy can be improved, it shows potential for eyes-off point-to-select interaction with no need to refer to a screen.

It is interesting that all three modes suffered from inaccuracies in the orientation of the pointing gesture. The results seem to suggest that participants struggled to point in the correct direction for targets further away, regardless of mode, which is a surprising result given the apparent simplicity of the pointing gesture. The most likely explanation for this angular inaccuracy is that an error that would be insignificant for nearer targets becomes magnified over larger distances, making accurate pointing difficult beyond a fairly close range.[3] Despite this, we expected much lower impacts on the orientation error at the relatively short distances used for this study. Given that the errors recorded were present over a large number of participants, this result suggests a possible design implication for future point-to-select systems. If pointing gestures can not be relied upon to allow accurate direct targeting of locations, then work is needed to help users accurately point toward the point-of-interest they wish to discover information about.

Previous point-to-select research (e.g., [48, 132]), and our aerial mode in this prototype, has used visual feedback to provide greater point-to-select accuracy. Similarly, more recent commercial systems (e.g., [85]) use a live camera view, augmented with geolocated content icons, minimising the orientation problem. But these methods still require the user to look at the screen rather than their surroundings. While our eyes-off mode was inaccurate for distance specification, it was no less accurate than

---

[3] A similar result was discussed by Strachan and Murray-Smith [143] around the same time as this research was originally published.

the other options for pointing. A solution that removes the need to tilt or specify distance, and instead offers feedback to help the user realise when they are pointing accurately at an object, could provide both eyes-off physically-grounded interaction, and a higher level of accuracy

As we saw in Section 2.3.1, previous work has successfully used non-visual modalities for location-based feedback. One clear possibility for eyes-off interaction is to provide vibrotactile feedback related to the density of geo-tagged content in an area. That is, as the user looks around an area, probing it with the device through pointing and scanning gestures, their mobile vibrates when they point to points of interest about which content is known to be available.

Taking this non-tilting tactile feedback approach removes the possibility for users to discover information about points of interest that are occluded, and makes it difficult to select objects that are partially obstructed by other objects in front of them (either physically or digitally) – a potential problem, particularly for areas with large quantities of geotagged content. However, in our previous work before this thesis we discovered that the majority of locations queried were very close to the user's location [120], and people rarely queried targets further away. Consequently, a sensible solution to this issue is to use eyes-off vibrotactile-supported target selection for objects that are close to the user, and fall back to traditional visual-based methods for targets that are further away. In the next section we investigate and explore the effectiveness of using this vibrotactile feedback approach to support discovery.

3.3 Investigating vibrotactile-supported pointing

From our initial study, it is clear that visual feedback can be slower and more cumbersome than eyes-off approaches, but designs with insufficient feedback suffer from fairly severe accuracy issues. Tilting to indicate distance does provide some accuracy in target selection, but the eyes-off benefit seems minimal, and only suitable for gross indications of distance. It is clear that in order to provide eyes-off discovery we need to consider alternative approaches.

Previous work has investigated using vibrotactile feedback to support mobile interaction, showing its benefits in many cases. Inspired by these systems, and with an aim to improve the accuracy and user experience of our previous prototype, we

Figure 3.8: The prototype in use: the SHAKE device (shown inset) vibrates when the user points toward a location with geo-tagged content available. Pressing the SHAKE's navigation button (shown at A) selects the target.

created a second discovery system to provide haptic feedback. Our aim in this second prototype is to help users discover and accurately point to digital objects associated with physical objects in their surroundings. In addition, we are particularly interested in whether the design can support usage while the user is moving. In this section we describe the design and evaluation of this second prototype, focusing on the accuracy that it can provide.

3.3.1 Prototype design

Our haptic prototype allows the user to discover geo-tagged information in the environment around them by making pointing and sweeping movements with a mobile device. When using our prototype, as the user moves around their environment, their position (latitude, longitude) is used to refresh a selection of available geotagged points of interest near their location. Holding the device in their hand, the user can point and scan around the points of interest nearby, feeling gentle vibrotactile feedback when they move their focus around each target, as illustrated in Fig. 3.8

The feedback provided helps to alert users when they point to a location about which geotagged content is available. The use of tactile feedback also allows users

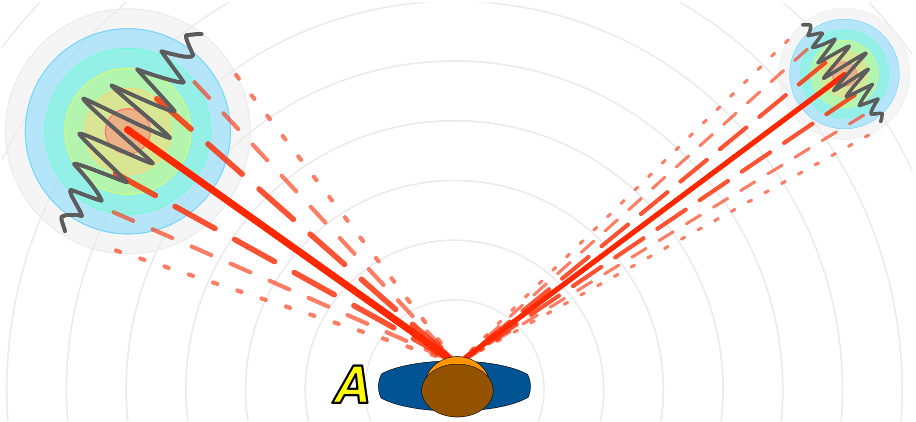


FIGURE 3.9: Vibrotactile feedback: the spread of the vibration area helps the user, at point A, to determine the size of the target (i.e., the quantity of information available), and the increase in vibration frequency toward the middle of the target guides them to its centre point.

to discover both the quantity of information available about a point and its centre point (see Fig. 3.9). Quantity is indicated by the spread of the vibration: targets with a large amount of geotagged content appear larger and take up more of the user's scanning range. To help the user determine the direction of the actual target that the content refers to, the vibrational feedback increases in intensity toward its centre. Importantly, unlike the approach taken in our initial prototype, there is no requirement to indicate the distance of the target – the maximum range is 100 m, and the system automatically selects the closest target if there are multiple options in the direction the user is pointing.

Our prototype system consists of a Sony VAIO Ultra-Mobile PC (UMPC), connected via Bluetooth, as in our first prototype, to a SHAKE motion sensor pack. For this design we used a UMPC rather than a PDA because, while the UMPC is larger, it has greater processing power and a larger, brighter screen for comparison with a visual mode of the system (detailed in Section 3.3.2). The SHAKE's tri-axis accelerometers, gyroscopes and magnetometers provide realtime orientation data, and its internal motor is used to provide vibrotactile feedback. The navigation button on its side is used to allow users to confirm a selected target. The SHAKE is attached firmly to the back of the UMPC so that any movements the user makes whilst holding it are directly recorded. All data received from the SHAKE, including timings, button presses by the user and logs of when vibrotactile feedback pulses were sent (and their intensities) are recorded by the UMPC at all times. An external GPS receiver provides location data.

Providing vibrotactile feedback on a device that is also used to gather sensor data potentially introduces problems with sensor interference. In order to address this, we first low-pass filter the accelerometer, magnetometer and gyroscope data, then combine these to obtain a more stabilised orientation matrix. The Euler angle for the heading obtained using this approach provides the system with a heading value that fluctuates less than using magnetometer data alone would allow. This technique introduces a small lag (approximately 200 ms), but we feel that this is acceptable to counter the more significant usability problems that would arise if we were to leave the sensor data unfiltered.

When the filtered compass heading intersects a nearby point of interest, the device generates vibrational feedback based on the tangential distance from the centre of the target, and increases the feedback intensity as the distance between the heading and target centre point decreases. When providing feedback, closer targets occlude those further away – we assume the user has an interest in the visible points of interest in their immediate vicinity.

While our earlier eyes-off design suffered from inaccuracy in target selection, the novel interaction method used in this prototype provides a simple way for people to accurately seek out digital resources, with feedback, whilst moving through their environment. We envisage this technique being used while both stationary and while moving, but believe it offers users the most benefits while walking and moving around their surroundings: the user can view, experience and interact with the physical features of their environment, while simultaneously exploring the digital accompaniments to the world around them, eyes-off.

In the next section we describe a user study undertaken to investigate the effectiveness and accuracy of the prototype, evaluating its performance against an alternative, visual-based system.

3.3.2 Experiment: evaluating vibrotactile selection while moving

When evaluating our initial prototype, we studied its usage in a experiment that looked solely at its accuracy while participants were standing still. When designing for engaging, eyes-off interactions, we argue that it is important that people are able to use the device while moving, to reduce its interference in their interaction with

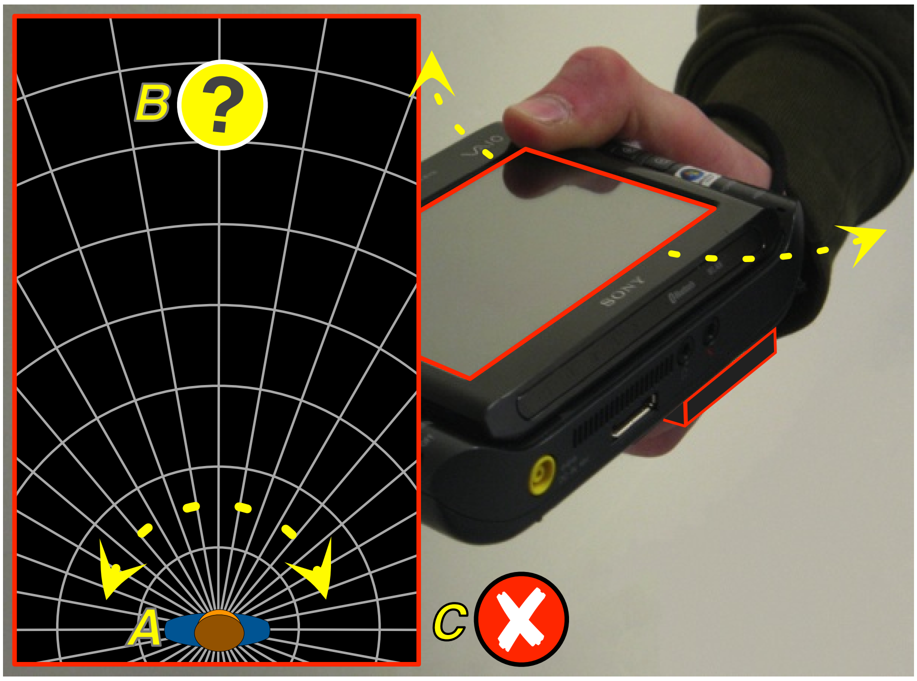


FIGURE 3.10: The visual system used for the study, showing (inset) the display from the user's perspective. The user is represented on the screen by the icon at *A* and a highlighted target is shown at point *B*. The icon used for targets not yet selected is shown at *C*.

the world around them. To evaluate our haptic system's usability and accuracy in such a scenario, we undertook a user study to investigate the system's usability when participants were walking. In this section we first describe an equivalent visual-based device that was constructed to allow comparisons between our prototype and a viable alternative system, then detail the method used to evaluate and compare these systems.

Screen-primary comparison system

To allow a fair comparison between the haptic system and a viable alternative we constructed a second prototype that is a screen-based analog of the haptic prototype, to the extent that it was possible. Where the haptic system provides vibrotactile feedback, the screen-based version shows icons as a visual representation of the points of interest that are available for discovery in the area around the user (see Fig. 3.10). Users hold exactly the same hardware as in the haptic system, but on the screen the device shows a radar-like display that rotates as the device is turned, updating in realtime to ensure that the display shows the targets that are currently in the user's field of view.

Feedback for target interaction is achieved by changing the colour of the target icons when the user points directly toward them, and the centre of each target is indicated by a cross (see points B and C, Fig. 3.10). In addition, the quantity of information available is indicated by the size of the icon: larger icons indicate more geo-tagged information is present.

Study design

Twenty participants aged from 18 to 55 were recruited for a 15 min laboratory study. Fifteen participants were students and five were members of university staff in areas unrelated to HCI. Ten participants were male and ten female. Six participants had previous experience of sensor-based interaction using a games console.

Prior to the study, participants were randomly allocated between two conditions: the haptic system and the visual comparison system. When allocating participants, those with prior sensor-based interaction experience were randomly divided so that an equal number used each system, then the remainder of the participants were allocated randomly to give ten participants per system.

Measures

Our main interest in this study was in evaluating whether haptic feedback could allow accurate point-of-interest selection whilst moving, comparing performance against the equivalent screen-based system. We measured user performance over several factors. specifically: whether users could find targets; the extent to which their normal walking speed was affected; the time taken to select targets; any false positives generated; and, participants' perceptions of the systems.

To quantify system performance we used the Percentage of Preferred Walking Speed (PPWS[4]) measure [111] as an indication of the system's effect on a user's normal behaviour, but used a slightly different method for recording this to mitigate concerns about its inaccuracy.

PPWS is measured by recording a participant's average preferred walking speed before the experiment, then expressing walking speeds recorded during the study

---

[4] See Chapter 2, Section 2.5.1 for a discussion of previous related research into mobile device usage while moving.

as a percentage of this value. However, during an initial pilot study between our systems, we found the commonly-used pedometer-based measurement of PPWS to be unreliable for shorter distances. When using the system in a pilot study, participants often took more steps over the same distance than they had when their pace was measured beforehand. Using this result to calculate a percentage of normal walking speed behaviour would show a speed improvement, when this is not actually the case.

To solve this accuracy problem, we opted to place a tape measure at the side of the walking circuit, and referred to this to measure the distance walked by participants, rather than using a pedometer. Measurements were taken from the front of the participant's foot to the nearest 10 cm. In addition, to correct for the curved walking course, at the end of the study all recorded distances were scaled up in line with a previously-calculated average maximum circuit length of 10.2 m, obtained prior to the study by measuring the average distance walked around the circuit of three non-participants.

In addition to the PPWS measure, we were interested in the time taken to select targets, which could indicate whether targets in certain positions are more difficult to select than others. We also recorded the number of false positives—that is, when a participant pressed the selection button when not actually pointing toward a target—allowing us to highlight interaction issues or cases of feedback confusion.

Finally, we used a post-study questionnaire based on the NASA TLX instrument [56], rating six aspects of participants' perceived performance: mental demand, physical demand, time pressure, performance, effort and frustration – from 1 (low demand) to 7 (high demand). We also recorded any comments given by participants, and noted interesting participant behaviours.

Tasks

We based our tasks on the adjusted PPWS methods used by Pirhonen et al. [113] to test a mobile device, in which participants walked around an indoor circuit, rather than over an outdoor course. Our aim was to simulate a realistic usage environment, but also to maintain control over the variance that can occur in studies such as these. During the study, participants walked around a circuit in a university corridor, as shown in Fig. 3.11, negotiating cones (representing obstacles) while using the system

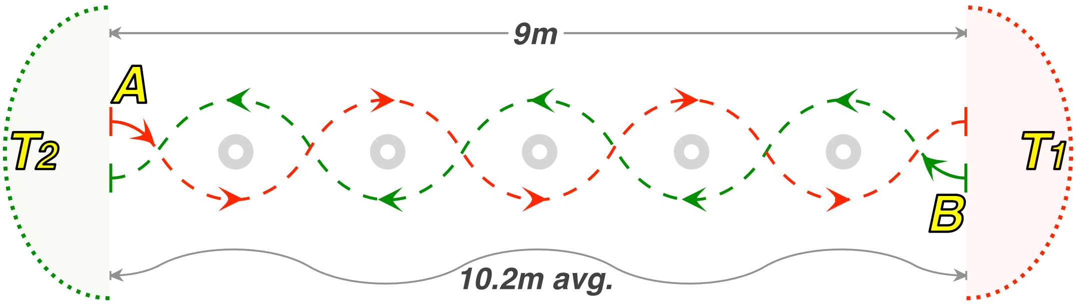


FIGURE 3.11: Study layout: grey circles indicate positions of cones; arcs T_1 and T_2 are target zones. Participants walk from point A to the end and repeat from B to return.

at the same time to select targets. The use of cones as obstacles in the circuit aimed to simulate the indirect walking behaviours that occur in real-world situations in response to obstacles or other pedestrians, but ensured that participants were not in any real danger of walking into anything. We also wanted to try to ensure that participants did not simply look at the device for the entire route (which could be the case for both systems on a simple circuit where little attention is required), in an attempt to simulate real-world eyes-off interaction.

Participants completed three primary tasks during the study:

1. Walk three 9 m lengths of the corridor to measure their average preferred walking speed.
2. Find and select each of 30 targets at opposite ends of the circuit (see Fig. 3.11):
   (a) Stand at point A, touch the device screen and start walking toward point B. While walking, attempt to select a target in zone T_1.
   (b) Repeat from point B to A, selecting a target in zone T_2.
   (c) Continue to repeat these steps, selecting 15 targets at each end of the course until all 30 targets have been selected.
3. Complete the TLX questionnaire to rate perceptions of the system.

The targets used were identical in size, and appeared 12° wide on each system. Targets were equally spread over zones T_1 and T_2 at the same relative positions in each zone, but participants could only see or feel feedback for one target at once. After a target had been selected, the next target was made visible. The order of the targets (see Fig. 3.12) was randomly generated before the study, but this order was the same for each participant and system.

Procedure

At the start of each study session participants were met individually and given an introduction to the system and its purpose via a verbal walkthrough of a typical usage scenario. Participants were then shown a demonstration of the system to select example targets in the lab. Following this, participants used the system themselves to select three sample points as a form of training.

Each participant was then taken to the corridor used for the experiment and, before starting the study tasks, was asked to walk three 9 m lengths of the corridor at their normal pace while the researcher timed them. This initial task allowed us to calculate each participant's average preferred walking speed. Participants then used the system to select targets while walking along the corridor circuit, attempting to perform both tasks at the same time.

Participants started at point *A* and were asked to simultaneously start walking and touch the device screen to begin target selection. Once walking, participants followed the line of the course, weaving around the cone obstacles while trying to select the current target. When the participant successfully selected the target, they stopped walking immediately so the researcher could note down the distance they had walked. Participants were told they should not stop walking until they had either selected the target or reached the far side of the course. If the participant reached the end of the course without selecting the target, they stopped and selected the target from the end of the circuit, but the researcher noted this event. After the participant successfully selected a target, they repeated the procedure from the opposite end of the corridor (point *B*), continuing to repeat this alternately until all targets had been selected, walking 30 lengths of the corridor in total.

After selecting all targets, participants completed the TLX questionnaire and offered any verbal feedback resulting from their usage of the system. Finally, at the end of the study all participants were rewarded with a bookstore gift voucher as a token of our appreciation.

3.3.3 Findings

During the study (one haptic, one visual, both students in the 18–25 age group, and with no prior accelerometer experience) found it very difficult to undertake the tasks they were set, and struggled to make sense of and complete the experiment. When initially analysing recorded data, the PPWS measurements from these participants were found to be significantly set apart from the rest (over two standard deviations from the mean). Consequently, we removed these two outlying participants' data (target selections for 60 targets, observations and their verbal feedback) before continuing with analysis of the results from the remaining 18 participants.

In addition, 20 cases (3.7 % of the remaining targets) where participants reached the end of the circuit without successfully selecting the target were removed from the dataset. Any PPWS measurements taken from these results would falsely show a speed decrease, and other measurements would not be representative, as participants were not moving when they eventually selected the target in these cases. No further outlying results were found, and we conducted detailed analysis on the data from the remaining participants and 520 total targets selected.

Participant performance

Table 3.2 shows the mean and standard deviation of each measurement for each system over all targets. Participants using the haptic system have achieved nearly 38 % of their normal walking speed, while participants using the visual system were able to walk slightly faster, at around 44 % of their preferred speed. Participants using the haptic system also walked comparable distances, and took a similar amount of time selecting each target overall, with fewer false positives generated per target.

Measurement (units)	Haptic		Visual	
PPWS (% of original speed)	37.7	(19.6)	43.6	(18.1)
Distance walked per target (m)	3.5	(2.5)	4.3	(2.8)
Time to select each target (s)	6.7	(4.8)	7.2	(4.5)
Nº false positives per target	0.5	(1.2)	1.2	(1.7)
Original walking speed (m/s)	1.5	(0.1)	1.4	(0.2)
Original walking speed (km/h)	5.4	(0.36)	5.0	(0.72)

TABLE 3.2: Means and standard deviations (in parentheses) of each of the measures recorded.

This is a promising result when compared to the large differences between visual and eyes-off modes in our first prototype, and suggests that the use of haptic feedback has improved the usability of the eyes-off version of the system. While statistical analysis using GLM ANOVA shows that a significantly higher overall PPWS is attained when using the visual system ($F = 10.25, p = 0.001$), the haptic system has generated significantly fewer false positives than the visual ($F = 29.5, p < 0.001$). No significant difference was found between the two systems in the distance walked or the time taken to select targets.

Regionally-separated targets

It is possible to merge data from the targets for both routes along the circuit and analyse them together, due to the fact that the targets are mirrored at each end of the course. For example, target 12 in zone T_1 was in the same position from the participants' perspective as target 27 in zone T_2, so their data can be combined into a single position for analysis. After pooling the data from these targets, we are also able to segment target positions into separate regions in order to highlight any interesting results that may have arisen due to the target locations. We opted to split the targets into left, centre and right of the user's position.

Figure 3.12 illustrates the mean PPWS per target and per region. Targets in zone T_2 are shown rotated to be in line with those in zone T_1 to allow visualisation of differences in performance at each relative target location. Each circular icon shows the order in which participants visited the targets. Outer bars show the mean PPWS per target, and inner bars combine ten targets to give a mean value for left, centre and right regions.

Interestingly, when segmented into these sections, it is clear that targets to the left and right of the participant have allowed them to maintain similar PPWS rates between systems. ANOVA analysis supports this, with no significant difference found between systems for targets to the left and right of the participant. Participants using the visual system were able to maintain a significantly higher PPWS for targets in the centre of their field of view ($F = 5.52, p = 0.02$), however.

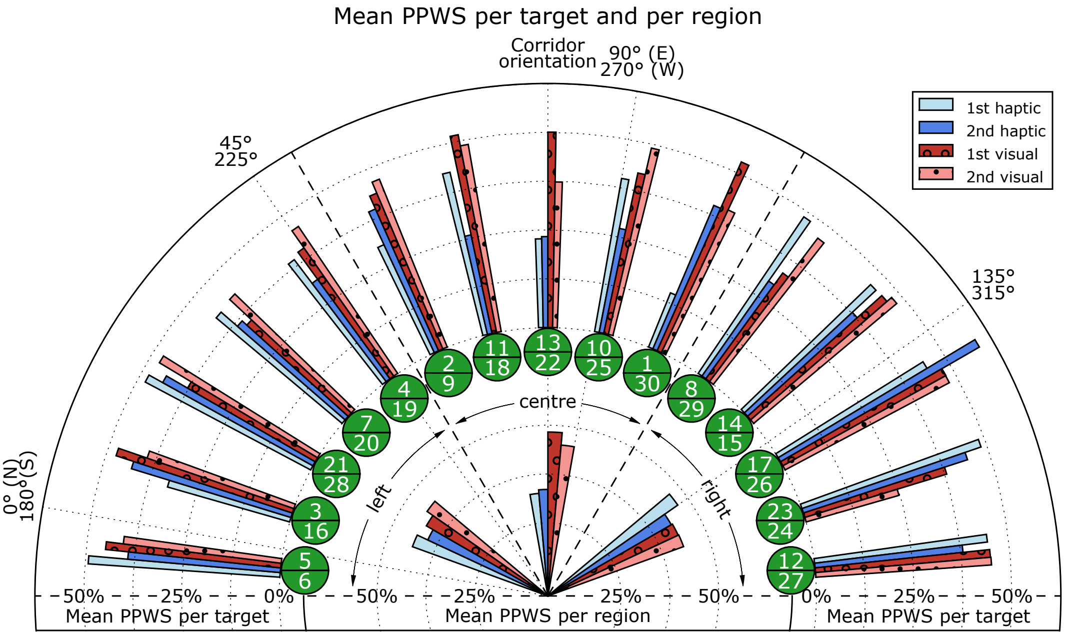


FIGURE 3.12: Regionally-separated targets: circular icons show the location of each target, and the order in which they were selected (numbers 1–30). Outer bars indicate the mean PPWS per target for each system, separated into the two attempts that participants had to select a target in this position. Inner bars combine groups of targets into regions positioned to the left, centre and right of the user: the visual system allows a higher PPWS for targets in the centre of the viewing region, but there is no such difference for targets to the left or right of the user.

False positives

This pattern is not shown when considering the false positives generated in each region. In fact, the haptic system actually generates significantly fewer false positives regardless of the target region (left: $F = 10.05, p = 0.002$; centre: $F = 7.05, p = 0.009$; right: $F = 13.49, p < 0.001$).

Looking closer at this result, we can compare the direction the participant was pointing when they triggered the false positive with the area displayed on the visual system's screen at the time. Seventy-five percent of the false positives recorded on the visual system were found to have been recorded when the target was visible on the screen but not selectable (i.e., not directly in front of the participant). Figure 3.13 illustrates the spread of false positives for each system, shown in relation to the target position, highlighting the denser concentration of false positives recorded on the visual system in the area where the target was visible.

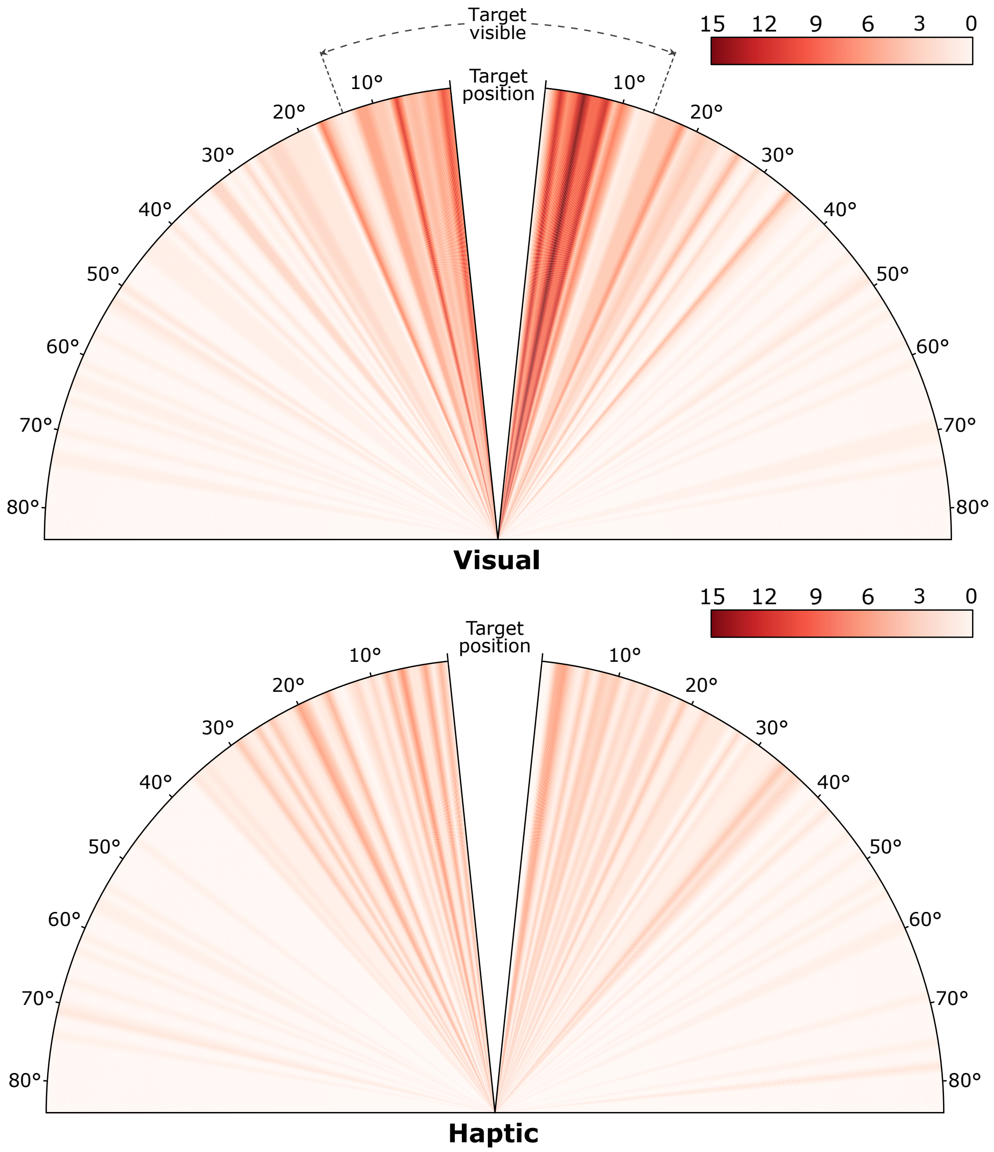


FIGURE 3.13: Spread of false positives recorded from each system, showing the number recorded in the area around each target.

Target searching behaviours

Participants' target searching strategies were visually classified into six different types of searching behaviour. Figure 3.14 illustrates examples of each type of behaviour. Table 3.3 shows the distribution of each type of behaviour for each system. The most common behaviours for each system were the 'directly to target' and 'probing around target' classifications, with around 70 % of all targets selected using this approach.

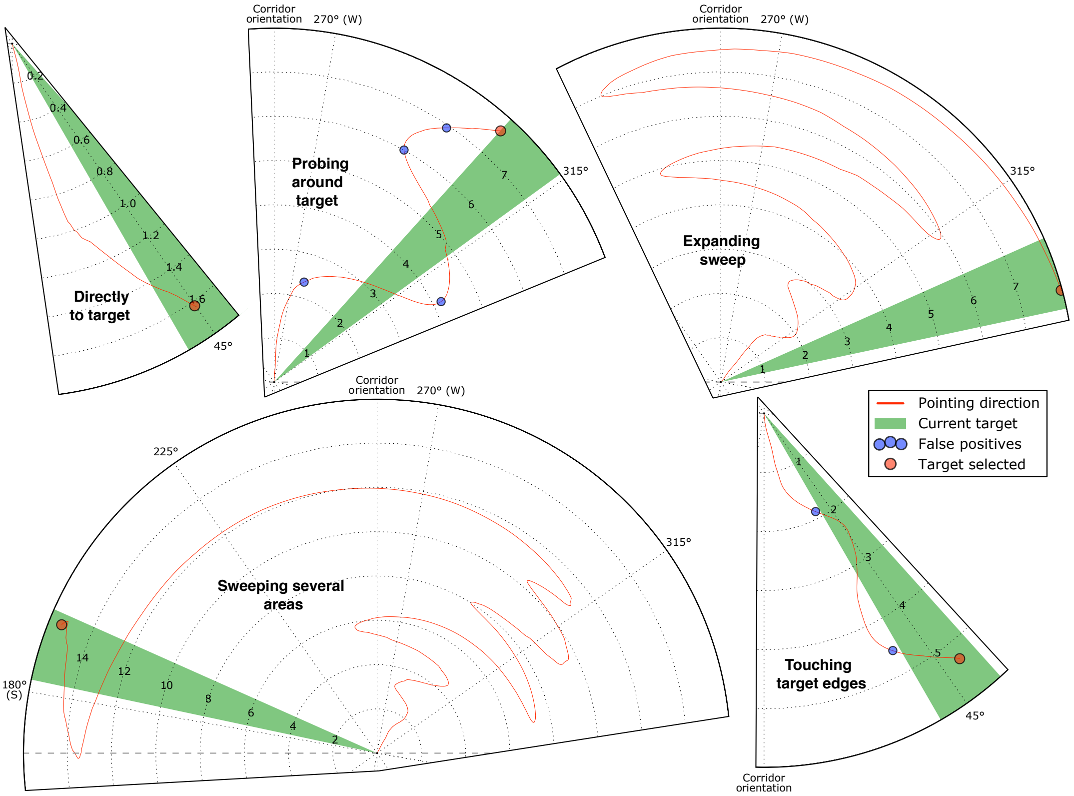


FIGURE 3.14: Representative target searching behaviours over both systems. The direction the participant was pointing is shown over time (s) from the tip to the outside edge of each arc. All arcs are oriented in the same direction relative to the orientation of the corridor participants walked down. The target the participant was attempting to select is shown by the shaded area. False positives, and the successful selection of the target are highlighted on each trajectory line.

Behaviour	Haptic (%)	Visual (%)
Directly to target	33	49
Probing around target	34	24
Expanding sweep	18	14
Sweeping several areas	6	12
Touching target edges	4	1
Other / unclassified	5	0

TABLE 3.3: Distribution of participants' target searching behaviours for each system, showing the percentage of targets that were selected using each strategy (illustrated in Fig. 3.14).

Participants' subjective ratings

Figure 3.15 shows the spread of each of the TLX ratings given by participants. ANOVA on these responses shows a significant difference between those participants using the haptic system and those using the visual system for their perception of the time pressure ($F = 6.12, p = 0.027$): participants felt the haptic system put them under less

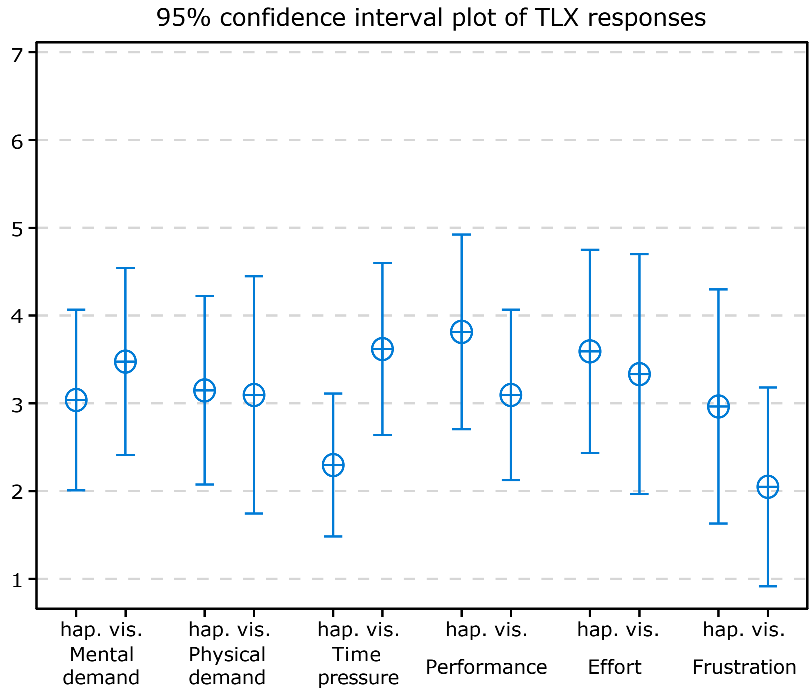


FIGURE 3.15: Spread of TLX responses for each system.

pressure to complete the task quickly. No further significant differences were found when considering any of the remaining TLX ratings for either system ($p > 0.05$).

Observed behaviours and verbal feedback

All participants initially chose to use their right hand to hold the device, and used their left hand to tap the screen to begin the task. After finding a small number of targets, five participants using the visual system began to use both hands to hold and steady the device, while all haptic participants continued to use only one hand to sweep and select.

The majority of participants using the haptic system tended to look almost straight ahead, appearing quite focused on the vibrational feedback they were searching for. While the haptic feedback did allow them to look away from the device, there was still a noticeable element of concentration required to interpret the feedback. Most visual participants appeared from observation to struggle to weave between the cones while looking at the screen, though only three mentioned this in post-study comments.

Feedback about the haptic system was broadly positive – one participant using the system commented: *"if you stop looking down at it it's much easier. I'd like to be able to use this in real life – it would be very helpful,"* while a second stated that *"if you're* *not looking at it you're concentrating much more on the feeling rather than the screen,"* and *"once you get the hang of it it's really easy."* Another participant stated that they *"didn't need to concentrate very much,"* but also that it was *"easy to go past too far"* and have to backtrack to find the target.

Participants using the visual system offered comments more aligned with tourist applications, such as: *"it would be really helpful for my travelling instead of going to tourist information,"* though several criticisms of the system were also raised: *"I have to catch [the targets] as they go past,"* and *"the delay [lag] makes it much harder."*

3.3.4 Discussion

Analysis of the results of this second study indicates, encouragingly, that the addition of haptic feedback has helped users to find and select targets much more effectively than our first eyes-off prototype earlier in this chapter. The use of haptic feedback has greatly improved the precision of the point-to-select interaction, without requiring a visual interface that could detract users from their physical surroundings.

Participants in this second study were also able to select targets while simultaneously walking and navigating around obstacles. Those using the haptic system were able to maintain walking at around 38 % of their normal speed, on average. Participants using the visual control system achieved a slightly higher percentage of their normal walking speed – around 44 % on average, but the difference between the visual and eyes-off systems is far smaller than in our first prototype.

When looking more closely at the results, we can see that for targets to the left and to the right of the user there is no significant difference between systems in the speed achieved, and only for targets in the centre does the visual system allow a significantly faster walking speed. This result is particularly positive, and highlights how the addition of simple directional vibrotactile feedback can improve targeting accuracy.

Throughout this study, participants slowed down their walking in all cases, as might be expected when concentrating on using a mobile device. However, there was no great gain in visual performance, and participants found all the targets using the haptic system. Although a comparison of targets overall (i.e., when not grouped into separate areas) shows that the visual system allows a significantly faster walking speed, it is important to consider this result in light of user interface familiarity. Users have likely had very little (if any) experience with vibrotactile feedback for interaction, other than as a background alert from a mobile device or as part of a games console, but most will have had extensive experience with visual, GUI-based systems. With further experience (and thus training) in haptic interaction, users' accuracy and ability to interact while walking would be expected to improve.

Also worth taking into account is the short delay in feedback between pointing to a target and feeling its vibration. This delay was caused by a range of factors, including the processing that is necessary to interpret inertial sensor data, Bluetooth round-trip communication times, and the startup time necessary for the vibration motor. It is possible that those behaviour types in Fig. 3.14 where participants repeatedly interacted with the feedback may have been caused by this short lag, but we expect this to become far less of an issue as tactile feedback response times decrease in future mobile devices. Future devices are also likely to be fully equipped with internal inertial sensors, lessening this delay even further and improving the response of our tactile feedback.

The similarity between haptic and visual results when targets are separated into regions should be seen as a positive point: with very little training, when interacting with targets to either side of their location, there is no evidence that the familiar, visual-based system provided participants with any additional benefit. In addition, while the visual system requires concentration on a screen, the haptic version allows the user to focus on their surroundings, as observed during the study – participants using the haptic system looked straight ahead or occasionally at the cones they had to walk around, rather than at the device.

For targets directly in front of the user, the visual system has allowed participants to walk significantly faster than the haptic system. This could be due to the ease of this task when targets are directly ahead – participants started the task and were able to see and select the target almost immediately, whereas participants using the haptic system may have automatically gone straight into a sweeping motion, after which more searching was needed to get back to the initial starting position. For areas to the left and right of the user, the haptic interface shows promise, but the design evidently needs improvements for targets in the area directly ahead.

Unexpectedly, the visual system has generated more false positives than the haptic version. Referring to the visual data (see Fig. 3.13) and the 75 % of false positives that occurred when the target was visible, this appears to be due to participants predicting target positions before they were selectable. When using the haptic system fewer false positives are generated, at the expense of time to select targets in some cases. This result highlights the difference in feedback resolution between the two systems, and the trade-offs that result from the differing response levels. Current visual systems often use similar interaction methods to that used in our visual display, allowing users to see potential targets' locations before they are able to select them. This can, however, result in unnecessary button presses. The haptic system tested here only notifies the user of a target's presence when it is selectable, resulting in fewer false positives, but a more exploratory discovery process.

When rating the systems, participants using the haptic prototype rated their perception of time pressure significantly better than those using the visual system, with no significant differences for any other ratings. Verbal feedback, too, seems to have indicated user appreciation of the haptic system, with several positive comments about its ease of use. Participants using the visual system offered several similar comments, but some found the interface difficult to use. Interestingly, only one of the haptic participants commented that they had difficulty using the system, despite the fact that the systems are essentially the same in their interaction method, and only the feedback differs. One explanation for this result is that participants using the visual system were able to see their errors, such as occasions where they missed targets and had to backtrack, or when they came close to selecting but pressed the selection button at the wrong time. Participants using the haptic system, however, could only detect these events on occasions where they skipped over the target and had to revisit it – pointing near to the edge of the target would not provide feedback.

3.4 Conclusions

In this chapter we began by exploring the extent to which screen-primary feedback is necessary for physically-grounded discovery of geotagged content. We then concentrated on improving the performance of an eyes off approach to discovery by augmenting a point-to-select gesture with vibrotactile feedback. Throughout this chapter we have demonstrated novel interaction methods, embodied in our design prototypes. As demonstrated in our two studies, these eyes-off techniques are usable without the need for people to focus their visual attention on the device.

In our first experiment, the eyes-off mode of our prototype allowed people to perform target selection quicker than that of a visual alternative. The study provided evidence to show that while screen-primary designs are likely to be far more accurate for this particular type of interaction, there are some benefits in the quicker interactions of eyes-off or minimally visual devices – users of our eyes-off prototype appreciated the low level of effort required to use the system to select targets. Pointing accuracy degraded with distance throughout our study, even when an aerial view was used to aid with precision. This is clearly an issue to address when considering the design of future point-to-select systems – if further eyes-off interaction methods are to be developed then targeting accuracy is essential to allow for the creation of usable devices.

Our initial designs used a novel two-phase point-to-select gesture, but results suggested that participants had some issues converting the view in front of them into a discrete value in metres, regardless of the system used. In our second experiment we investigated the viability of using vibrotactile feedback to augment and improve the accuracy of point-to-select gestures. This study was designed to test the interaction in a more realistic usage scenario than our initial study – rather than investigating the system with participants standing still in a lab, we conducted the trial while participants were walking around a simple obstacle course. The results of this experiment showed a large improvement in the performance of the eyes-off prototype against our first design, with participants accurately discovering targets in a wide range of positions. In addition, despite reported unfamiliarity with pointing-based vibrotactile interaction, in two-thirds of cases participants were able to maintain the same proportional walking speed as those using a visual-primary version.

In reality, of course, obstacle avoidance is not the only element of interaction with the physical world – interaction and communication with other people, crossing roads or simply observing the scenery are just a few of the many things people do while moving around their environment. However, we argue that these additional tasks support our argument for a need for eyes-off interactive devices that allow usage in parallel with users' everyday tasks. This was demonstrated in the visual case during our second study, where exploration of the information space required a 'heads-down' interaction style which appeared to be distracting to the users. Haptic feedback allowed both interaction with the environment and discovery of information in an eyes-off manner.

3.4.1 Designing for eyes-off, physically-grounded discovering

From the design and evaluation of the prototypes in this chapter we can extract several important design properties and recommendations for future devices that explore this design space. These will be refined and extended in subsequent chapters as we investigate each of our elements of eyes-off physically-grounded interaction.

It is clear from the experiments conducted in this chapter that, while screen-based systems are likely to continue to offer more accuracy, eyes-off designs, when carefully tailored to the task at hand, can support complex interaction. The accuracy of eyes-off systems is likely to be influenced by several factors, including target distance, the use (or absence of) non-visual feedback, and the target's position relative to the user. Previous work has shown the popularity of pointing to select targets, but it is evident from our results that accurately targeting distant locations, even with a screen-primary system, may be problematic. Broad 'near,' 'middle' and 'far' selections can be achieved by using tilting to refine a gesture, but for greater precision further feedback is required. Minimal visual feedback can allow users to explore and discover with slightly greater accuracy, but the difficulty of conceptualising a given distance value and relating this to a real physical distance is likely not worth the extra mental effort. Instead, the use of tactile feedback can help users achieve accurate pointing, at the cost of being able to easily specify distance.

Pointing Pointing as an interaction method can be accurate, but the gestures used should be straightforward. Specifying distance in the same pointing interaction only serves to make both gestures less accurate.

Point-to-select gestures are most accurate at short distances, due to minor errors becoming magnified when users attempt to point to objects further away. Accuracy decreases as the target distance increases.

Accuracy To improve the accuracy of eyes-off pointing gestures, feedback should be used where possible. Without on-screen, tactile or audio feedback, eyes-off designs can offer only broad measures of accuracy.

With feedback, precision is greatly improved. In some cases, particularly when targets are not in front of the user, tactile feedback offers an improved experience, with fewer false positives, and no evidence of a penalty in walking speed.

Distance Distance specification may initially seem like an attractive way of increasing selection accuracy, but it is difficult for users to convert observed physical distances into a system input. Furthermore, it is likely that distance specification is not necessary for many of the most common targeting tasks involving nearby and clearly visible objects.

CHAPTER FOUR

Displaying

In this chapter we move from investigating eyes-off discovery of geotagged digital content to look at how this information might be *displayed* in a similar manner. We investigate this initially via the use of personal projectors, then subsequently show how vibrotactile feedback can be used to give a broad indication of content types in situ, without screen-focused interactions. We then turn to consider how tactile feedback might be used to display the location of *dynamic* location-aware elements in the user's vicinity, rather than just static content. Our approach in this chapter is again to focus on the evaluation of novel prototypes, demonstrating how the displaying of geolocated content need not focus entirely on a screen. The results from three separate experiments to evaluate these prototypes help to refine and extend our design properties for future physically-grounded mobile devices.

4.1 Introduction

Previous research has shown how using a screen-primary mobile device can lead to a loss of wider focus on all but the most important events [65]. As we saw in Chapter 3, systems that support non-screen-primary interactions can potentially allow people to focus on the environment rather than the device. Our earlier evaluation of haptic feedback while participants were walking showed that an eyes-off design was no less effective than a visual system in two-thirds of cases, and, while using vibrotactile feedback, people did not have to look at the device to discover content around them. It is clear that we have been able to address this issue to some extent.

But the discovery of information is only the first step of the process of location-based interaction. In this chapter we explore whether it is also possible to physically ground the *displaying* of geolocated content. Typically, when using a screen-focused device, there is no break between the discovery and display of content – a screen can be used for both parts of the process. When designing for eyes-off interaction, however, this is potentially a more difficult proposition. While a screen can interfere with a user's immersion in their environment, it does clearly provide a quick and easy way to browse. Our approach to this issue is to extend the use of the mobile as a pointing device for physically-grounded interaction, adopting as a foundation the directional tactile feedback successfully demonstrated in the previous chapter. Here, however, we investigate whether the initial discovery gesture can be augmented in other ways to allow people to browse the content itself eyes-off.

Clearly, in order for certain types of digital content to be displayed—images and videos, for example—a display of some sort is required. We begin this chapter, therefore, by investigating whether a visual, but non-screen, approach can support the sorts of physically-grounded interactions we aim to achieve. With our first prototype we consider whether handheld projectors might be practical for in situ displaying of the content located by vibrotactile-supported discovery.

Informal feedback from users of our initial prototype is promising, but current hardware issues—specifically the poor visibility and low brightness of current pico projectors—limit its potential in the sorts of scenarios we imagine. Consequently, our second prototype turns to consider whether tactile feedback could be extended in depth to provide a second layer of interaction for displaying broad categories of content. Its simple additional gestures, designed with the findings about the complexity of our earlier point-and-tilt gesture in mind, give the user feedback about the presence of different content types. We evaluate this prototype in two separate studies, and the results show promise for future hierarchical tactile interfaces.

Feedback from participants in these studies hints at a different area of potential for tactile feedback, however – particularly for mobile tasks that are perhaps even more grounded in the physical world than the discovery and display of geotagged digital content. In the final section of this chapter, then, we move away from direct displaying of discovered information to consider whether there are benefits in displaying *dynamic* content – geolocated elements that may be repositioned at any time.

The most common dynamic elements of any geolocated scenario are people themselves – their locations can be constantly updated depending on their movements through a space. Our third prototype investigates whether displaying a broad representation of a group's dynamic centre point to each individual member can help them find each other efficiently. We evaluate this prototype in a user study conducted over a large area, and the results demonstrate its benefits, showing how it can support dynamic, eyes-off group meetups. The refinement and evaluation of this and our previous prototypes help to improve and extend our design recommendations for eyes-off physically-grounded interaction.

4.1.1 Physically-grounded displaying

The following scenario illustrates how Alex—who used his mobile in Chapter 3 to discover geotagged content eyes-off—might use his device to filter, browse or display the content and people around him in a more physically-grounded manner.

> *Walking around town, Alex holds his mobile phone in hand. He points again to the music store over the road: the device vibrates, telling him that there's new content from artists he might like. He pauses, and, after another simple gesture, feels vibration feedback letting him know that there are new tracks available. Alex presses a button on his device and begins to listen to a preview of the new album. While he walks and listens, Alex casually points towards an event banner on the wall next to him to see what the show is about. His mobile vibrates, and its internal pico projector turns on to offer him a trailer. As he's pointing and browsing around several projected trailers to view the previews, the phone alerts him that his friends Daniel and Lucy have sent a meetup invitation. Alex briefly scans around with his phone until he feels the distinctive homing vibration. He starts walking in the general direction of the meeting point, and happens to see Lucy while he's on his way. The friends walk and chat together, both easily able to keep track of the haptic feedback at the same time, knowing that it will lead them to Daniel . . .*

This scenario illustrates three key behaviours that are explored in this chapter. Alex is using his mobile device to project content related to the things he discovers, and this content is displayed directly on the objects queried. He is also able to display content types in a broader eyes-off way, using additional vibrotactile-assisted filtering gestures to determine whether any of the types of content he is searching for exist within the target he has pointed toward. Finally, Alex and his friends are able to use a vibrotactile 'beacon' to display a dynamic mutually-convenient and constantly-updating meeting point, which helps them rendezvous.

The increasing availability of handheld (pico) projectors has opened up possibilities for lightweight projection-augmented interaction, stimulating research into many new devices and techniques. As might be expected, a great deal of this research, particularly that in the mobile domain, has concentrated on augmenting real-world objects with projected features (e.g., [91, 99, 146]). However, there is a lack of investigation into how people might make the transition between these physical and digital aspects – that is, moving from discovery of digital information to displaying the content itself. Previous work typically assumes a seamless flow from mobile interaction to projection, but often does not consider how the step from real-world to projected content might actually take place.

The processes of discovering and displaying geolocated information are often split between the physical and digital domains, with a choice between either a real world location-based object (an event poster outside a venue, for example) or its digital counterpart (the venue's webpage viewed on a mobile device, or later on a PC). We argue that combining vibrotactile discovery with personal projection might be able to support a smoother user experience by displaying digital content in situ. This transition between discovering and displaying pairs digital content with the physical object that it augments, allowing the user to see the information they request projected onto the real-world elements it describes.

Our first prototype system aims to support this interaction via peephole-style gesture-controlled pico projection similar to that of Rapp et al. [115]. The design uses tactile feedback for discovery, and projection to display the resulting content. The sensor-based detection of motion that we use for movement and skew correction requires none of the image processing that similar camera-based methods (e.g., [7, 99]) depend upon, providing a subset of their functionality with the benefit that the user need not wear coloured finger tags or lanyards.

Hardware limitations of pico projectors mean that real-world handheld projection is impractical at present. But the use of tactile feedback to support a lower-resolution display of content type categories (rather than the content itself) offers an alternative approach to allowing eyes-off display of digital information. Our second prototype is inspired by previous research into structured audio hierarchies [16], but here we use tactile feedback to allow people to feel the presence of different types of content. The vibrotactile hierarchy we developed allows users to scan their environment with broad sweeping gestures to feel for content around them, then focus in on any particular element to 'zoom' into and deconstruct it into individual information categories. We do not aim to support every displaying scenario – it is likely that users will often wish to view images or videos immediately, for example. Instead we allow users to explore when desired – once zoomed into a location, our system provides different haptic patterns based on the types of content available. This is a similar approach to that taken by Luk et al. [94], who used piezoelectric actuators to provide a range of tactile patterns. Here, though, our approach is to use both patterns and gestures in an aim to provide straightforward feedback rather than requiring users to learn a vocabulary of tactile responses.

So far in this thesis we have focused on realtime eyes-off discovery and displaying, but evaluations of our second prototype in this chapter suggest that physically-grounded mobile interaction might be more effective when focused on slower, more reflective searching. Previous research has discussed the temporal aspect of mobile information seeking [17, 73], arguing for systems that support delayed, 'laid-back' searching, rather than immediate retrieval and presentation of results. For immediate information displaying it seems sensible to use a screen to display content in situ, and to use tactile feedback where appropriate via a delayed interaction style.

However, while screens might be more effective for much of the current visually-oriented geolocated content that is available, for things that are not fixed in a particular place—people, for example—there are further possibilities for physically-grounded displaying. Pedestrians are inherently dynamic elements of any geolocated scenario, being free to move at will throughout the space. Previous work has investigated how the locations of groups of people can be displayed visually (e.g., [106]), or how proximity-awareness can be used to give feedback when friends are close to each other or leave a group (e.g., [104]). Rather than focusing on extending our current tactile displaying interaction further, then, in the last part of this chapter we evaluate tactile feedback for displaying the locations of other people as dynamic geolocated elements.[1]

---

[1] The software for this prototype was predominantly created by our project partners, but refined and evaluated here. See Appendix A ([P6]) for a summary of the author's contribution to this work.

We investigate whether group location awareness can be provided in an eyes-off manner, removing the requirement for pedestrians to follow navigation directions, and instead displaying the locations of other people – not directly, but via a group central point. Our work is similar in some ways to previous tactile navigation systems (e.g., [43, 44, 70]), which used belt-based directional haptic feedback to help users navigate or avoid obstacles. In our approach, however, we use a handheld device for navigational feedback. Lin et al. [89], found that users were able to pay attention to their environment at the same time as using a tactile system, but used a Wizard-of-Oz approach to create structured tactons to guide users. In our approach we use realtime vibrotactile feedback from a handheld device to allow users to search for the feedback zone eyes-off and, further, only experience feedback when they feel it is necessary.

4.2 Personal projection for physically grounded displaying

In order to investigate the potential for projection in physically-grounded interaction, we developed a prototype that extends the vibrotactile discovery method from our earlier tactile design, allowing users to display discovered content in situ. Our design uses a Nokia N95 mobile phone attached to an Optoma PK-101 pico projector. Orientation and movement data are provided by a SHAKE[2] motion sensor pack firmly attached to the underside of the prototype.

Figure 4.1 illustrates the interaction flow of the device. The vibrotactile element of the system is the same as our previous design – information discovery is achieved by simple scanning movements while holding the device, with the user searching their environment to discover areas via tactile feedback. When feedback is felt, pressing a button ‘zooms’ into the requested element, allowing the user to display its content.

Content is projected where the user is pointing, as illustrated in Fig. 4.2, and is clustered into three categories (images, videos and webpages). Projections from the system use the device’s orientation to move the projected items to correct for users’ movement – a projected peephole similar to that of Rapp et al. [115]. The aim is to give users the impression that the content they are interacting with is attached to the surface they are projecting onto. Displayed objects appear to stay in fixed positions as the device is moved, allowing the user to point at individual items to select them, or

---

[2] See Chapter 2, Section 2.5 for more details.

Photos by anirudhkoul (Flickr)

FIGURE 4.1: Interaction with our prototype system. Left: Users browse their environment while on the move, discovering location-specific information via vibrotactile feedback. Centre: Once discovered, projection is activated, allowing in situ display of content. Right: At any point, users can tilt their phone back to capture a copy of the selected item on their mobile device.

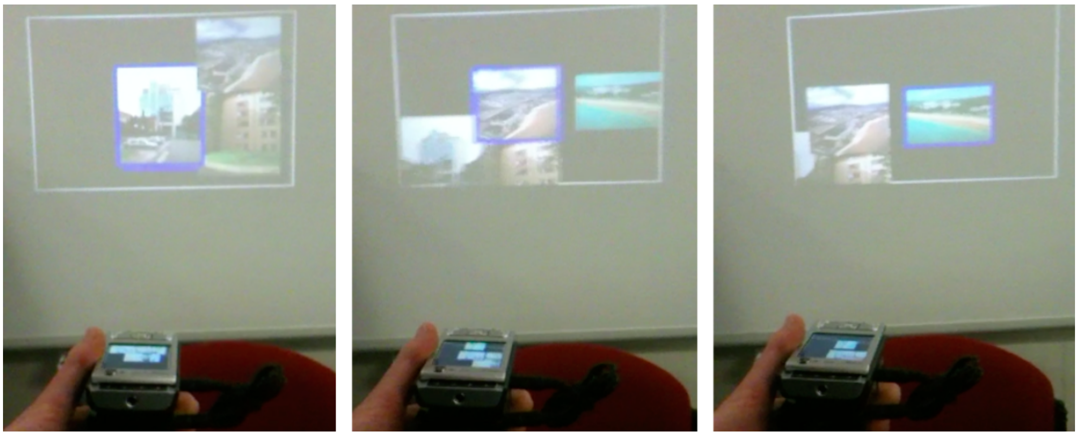


FIGURE 4.2: The prototype in use, showing a user selecting content from a cluster of images by pointing the device itself. Sensor data is used to correct for device movements and image skew to allow selection of individual items by moving the device, rather than a pointer. The current selection is highlighted by the system in each image.

around the current projection surface to see more of each content type. This allows quick and easy selection and displaying of individual content items without the need for a visual pointer, or indeed any method for touch-based interaction; instead, the user points using the projector itself.

4.2.1 Pico projection challenges

While our design shows how this type of interaction might work in future, in reality the limitations of current pico projectors restrict its current usability and future impact. The process of developing this early projection prototype highlighted several crucial issues with personal projectors – most notably their insufficient brightness, leading to poor visibility of the projected image. While there are many inherent challenges for in situ projection, both social and technological, the issue of brightness is critical – it is almost impossible to see current pico projected content in any level of sunlight.[3]

Because of these issues we chose to conduct only informal piloting of the prototype, as pico projections are currently inappropriate for the types of physically-grounded scenarios we envisage. The prototype was demonstrated to approximately 40 students during a lecture, and used by five visitors to a university open day. Responses from users and observers were focused upon the serendipitous aspect of the system, with comments such as *"it's like StumbleUpon[4] for the world around you,"* referring to the unknown aspect of the content that might be displayed.

Possibilities for content manipulation

Despite the limitations of current pico projection hardware, during the development and informal evaluation of this prototype several possibilities for content display and manipulation that might be possible without the need for projection were highlighted. Our pico projector prototype uses the projection device itself as a pointer to allow users to interact with the content they are browsing via a peephole-type approach. For closer objects the touch-based methods that are now common on mobile devices could offer simpler, more familiar interaction (pinch zoom; touch scrolling, for example – subsequently demonstrated with pico projectors by Cowan and Li [32]). But for surfaces that are too distant to touch or manipulate directly, or to reach via pico projection, a looser version of the peephole approach might be more appropriate .

For objects that are further away, our existing use of tactile feedback for discovery hints at a way to support a similar, but lower-resolution peephole-like interaction for displaying, without the need for screen-primary interaction. Rather than displaying the actual content immediately, we can use further gestures and tactile feedback to allow users to 'display' and filter the requested content types to some extent. In the next sections we discuss the development of a prototype designed to support this type of interaction.

---

[3] The Optoma PK-101 used in our prototype was the brightest commercially-available pico projector when the prototype was originally developed – its brightness is estimated by the manufacturer at 11 lm. Figure 4.2 shows the PK-101 used in a darkened room; in natural or artificial light the projected image is invisible. At the time of writing this thesis the brightest commercially-available pico projector is rated at 80 lm, but the device is considerably larger in physical size than the PK-101 we used and, despite the increased brightness, the image is still washed out and often near-invisible in natural light.

[4] Browser add-on for content recommendation; see stumbleupon.com.

4.3 Investigating vibrotactile-supported displaying

While our pico projector prototype shows promise, current hardware limitations make it impractical for real-world usage. However, there are possibilities for providing a low-resolution version of this peephole-type interaction by extending our haptic interface to support a multi-level tactile hierarchy. A similar approach was taken by Brewster et al. [16], who used structured audio messages (rather than tactile feedback) to help users navigate through four levels of an hierarchical menu, finding over 80 % accuracy for position identification. A vibrotactile approach to this type of interaction could allow users to 'display' the types of content available in a lower-resolution manner than screen-based interaction, but without the need to look at a screen or to wear headphones for audio.

We built a prototype to explore this interaction. The system helps the user to feel the presence of information in the space around them, based upon the same foundation approach as our earlier haptic discovering system, but extending this to allow displaying of content types. The system provides haptic feedback for discovery, and the physical area the information relates to and the amount of content available are indicated by the apparent size and position of the haptic pulses. By sweeping the device around a location the user can assess the possibilities. Once the user has selected a target, the system further extends the haptic response to help them display the information space in more detail, allowing them to zoom in on and filter the available data into different categories for exploration.

4.3.1 Prototype design

The *Sweep-Shake* system allows users to discover geo-tagged information in the environment around them with simple pointing and sweeping gestures, initially in the same way as our earlier haptic design in Chapter 3. This new approach uses the same UMPC, GPS receiver and SHAKE hardware as the original prototype, but in this design we add a second level of interaction, allowing users to 'display' content categories via tactile feedback.

As the user scans to browse the information around them they are able to leave the discovering mode and focus more specifically on one particular location, zooming in

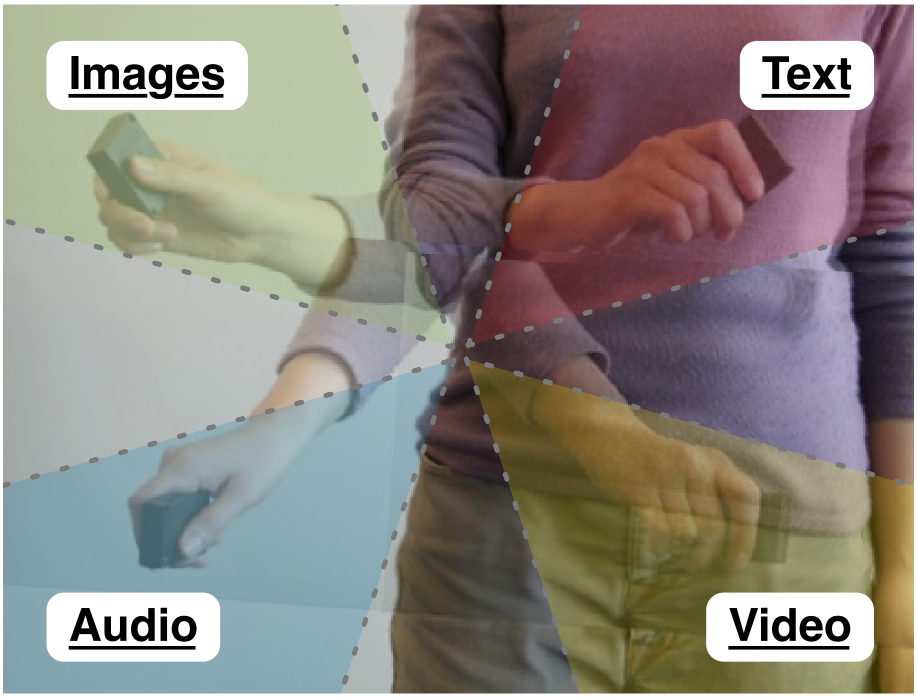


Figure 4.3: Haptic displaying: pointing toward the four haptic areas (shaded) triggers feedback. Haptic areas are 70° in width, and are centred around 45°, 135°, 225° and 315° from vertical.

to explore in more detail. In this prototype, pressing the SHAKE's navigation button whilst vibrational feedback is being felt causes the system to move into a displaying mode, and also causes a distinct 'zoom in' vibrotactile pulse to be generated to indicate to the user that they are focusing on a particular information hotspot.

Vibrotactile displaying

The system's displaying mode is intended to help the user to browse the available information about the place they have selected. However, instead of simply retrieving and displaying all nearby geo-tagged data, any relevant data is first segregated into four distinct clusters based on the types of information that exist at that place. Four simple gestures in this mode allow the user to request text, audio, image or video content related to their current focus. Mindful of the complexity reported by users of our earlier point-and-tilt prototype, we used a less complex four-way directional gesture, with no distance element. This novel interaction method provides a straightforward way for users to find and filter these specific types of location-related information.

By tilting the SHAKE to four corners around the location of their original target, the user can detect the presence of each of the four categories of content via vibrotactile feedback (see shaded zones in Fig. 4.3). At each of the four zones a vibrotactile pulse is generated to indicate the presence of that type of information, with a different pre-set pulse being generated for each of the information categories in an attempt to further improve the ease of use of the system. If no pulse is felt then no data of the requested type exists, and the user can simply zoom out (by pressing the navigation button again) and move on.

Whilst pointing to a specific category, if the user feels vibration to indicate that there is content of the type they are searching for then they can press the SHAKE's navigation button to retrieve and show it on the UMPC's screen, again feeling a distinct haptic pulse to confirm this action. The data is filtered before being retrieved to match the user's request: for example, if the user pressed the selection button whilst feeling vibration indicating the presence of images at a location, only images would be retrieved, and so on.

When the user has finished browsing the information, pressing the navigation button will return to the displaying mode, again generating a confirmation pulse, and the user can explore the other categories of information that are available. From this mode, pressing the button again will generate a distinct 'zoom out' pulse and return to the original browsing mode.

4.3.2 Experiment: exploring vibrotactile displaying while moving

We carried out a small field trial to examine the usage of the Sweep-Shake system in a realistic scenario, and to help identify aspects of its design that could be refined. This campus-based study made full use of the system's potential for finding information whilst mobile, using live information about several locations.

Method

Four participants aged from 18 to 35 were recruited for a 45 min field study. Three participants were students, one was a member of university staff; two participants were male and two female. One participant had previous experience of accelerometer-based interaction methods with a games console; the remainder reported no prior experience of this style of interaction.

Before the study commenced, we generated geotagged content hot-spots over five buildings located within the university campus. These target points all contained a minimum of five separate items of information about the building in question, ranging from web pages (categorised as text content), current interior and historical exterior images, to videos and audio recordings about recent and upcoming events. The system was in range of at least three of these information hot-spots from any point on the main route through the campus, giving participants a choice of several targets to discover and display at all times regardless of where they chose to walk.

At the start of each study session each participant was met individually and given an introduction to the equipment and its purpose, followed by a walkthrough of a short example usage scenario. Participants were then given a demonstration of the Sweep-Shake system and, as a form of training, were asked to practise both discovery and displaying of content in turn, finding and browsing up to three training targets from a window in our laboratory. These simulated targets were different to the five used for the study itself.

Each participant was then taken outside to the far edge of the university campus, and was asked to make their way through the area, exploring the space around them to find any information that might be of interest. As the participants moved through the campus the researcher observed from a distance, but did not interact with the participant. When participants reached the opposite side of the campus, they were asked to provide verbal feedback resulting from their experience of using the system, in a semi-structured interview. Participants were then given a bookstore gift voucher as a token of our appreciation.

4.3.3 Findings

Each participant chose to begin the study by standing still and systematically scanning around the starting location for potential points of interest. From this point onward, participants slowly walked through the study area, at first pausing whenever they felt a vibration, but later, toward the end of the session, continuing to walk and instead interpreting the feedback as they moved. Two participants held the device by their waist as they walked, moving it around to scan constantly, and then temporarily walking more slowly when they felt a vibration. The other two participants held the device in front of them, making intermittent but deliberate scanning motions.

All participants found each of the five targets and each participant displayed at least one category of content from each of the five targets available. Each participant explored each category of information at least once, but after this first interaction displayed interest primarily for text and photos, with only one participant viewing further video and audio content after they had initially discovered and displayed media from this category.

Verbal feedback

All participants commented that the system was fun to use, and that they enjoyed being able to point toward buildings to feel the presence of related information. Each participant also said that once they had been able to understand the system from their interaction with the initial target they discovered, they were much more confident in its usage. One participant felt that the interaction was like *"playing a game to catch the points,"* and that had the vibrotactile effect been more stable their performance would have been better.

Two participants recalled specific experiences of being lost in foreign cities where they would have liked to have been able to find information around them to help get a sense of their location. Two of the participants suggested a 'guide me' mode, similar to that of a standard GPS device but instead using haptic feedback to indicate the general direction of the target location. Two participants commented that they had felt the haptic feedback to be either off or on, rather than a steady increase in intensity toward the centre of the target.

4.3.4 Discussion

Encouragingly, the four participants in this initial trial were able to discover and display content from each of the available targets while moving, and after only basic training. Participants offered positive comments about the ability to discover, display and explore geo-tagged content, and enjoyed interacting with their surroundings in this physically-grounded way.

Participants did not seem interested in the audio and video content that was available, instead preferring to perform a quick scan of static text and images. This behaviour only became evident during reviews of the data after the trial, however, so we were unable to ask participants why they preferred these particular content types. While it is possible that this finding could be due to the small amount of content available for this study, it could also suggest that some users are reluctant to commit to watching or listening to discovered media of these types whilst on the move. Previous work has argued that it is important to design systems that support gathering content in a delayed manner, rather than immediately [73] – adding support for delayed interaction to the Sweep-Shake prototype could allow users to take a copy of interesting location information with them on their onward journey, possibly improving interaction with these types of more time-consuming content.

It was interesting that two participants held the device by their waist and used it almost as a background cue to let them know when interesting locations were available. This behaviour is an intriguing middle ground between traditional pull or push methods of interacting [26]. Users chose to combine pull (requesting information) and push (information automatically displayed) techniques by scanning in the background to create their own browsing experience, rather than requesting information about specific targets. This observation suggests that the system was successful to some extent in its aim to provide eyes-off interaction that avoids interfering in participants' normal behaviour. However, there were comments from some participants suggesting that they were not able to accurately interpret the tactile feedback to its full extent. The small number of participants in this study means that its findings are limited, serving mainly to demonstrate the usability of the system in a realistic environment. To examine users' behaviours and the benefits of the system for eyes-off displaying, we undertook a more comprehensive evaluation of the system, as detailed in the next section.

4.4 Evaluating vibrotactile displaying against a visual alternative

Our focus in the initial exploratory study of the Sweep-Shake system was on whether participants could discover and display content via vibrotactile feedback. The small number of participants in the study were indeed able to successfully use the system. However, it is important to also explore whether the system offers real benefits when compared to existing screen-focused approaches. To measure this, we undertook a second study to evaluate and compare the Sweep-Shake prototype against a visual alternative. This study focused on investigating users' understanding of the vibrotactile hierarchy, and its effectiveness with regards to accuracy and time taken.

4.4.1 Experiment: measuring the effectiveness of tactile displaying

In this second study we investigated the tactile displaying of content categories used in the Sweep-Shake system compared against an alternative, screen-primary system. Effectiveness was measured via participants' performance in discovering and displaying six targets located on objects around an outdoor study area.

Sweep-Shake system modifications

In order to support our interest in the effectiveness and efficiency of the browsing and displaying actions, the Sweep-Shake system was modified to focus on haptic discovery and displaying rather than viewing of the information retrieved. To enable this, we removed the capability to actually view the images, text, audio and video that were retrieved, aiming to ensure that users' behaviours were not affected by the quality or relevance of the content they saw. Instead, when a user discovered a target and zoomed in to display the four categories of content, the system offered only the haptic feedback indicating the presence of content, rather than allowing users to view the content itself.

In addition, to allow us to concentrate on participants' ability to find specific locations, each target became un-selectable after it had been selected once. That is, once a target had been discovered and each category of content within it had been triggered, pointing toward this target again no-longer triggered vibration feedback. Finally we attached the SHAKE sensor pack to the back of the UMPC (instead of it being held separately) to ensure the same interaction movements as with the screen-primary comparison system detailed below.

Alternative, screen-primary design

To help evaluate the prototype we constructed a comparison system using a screen-primary interface instead of tactile feedback. This system is a visual analog of the

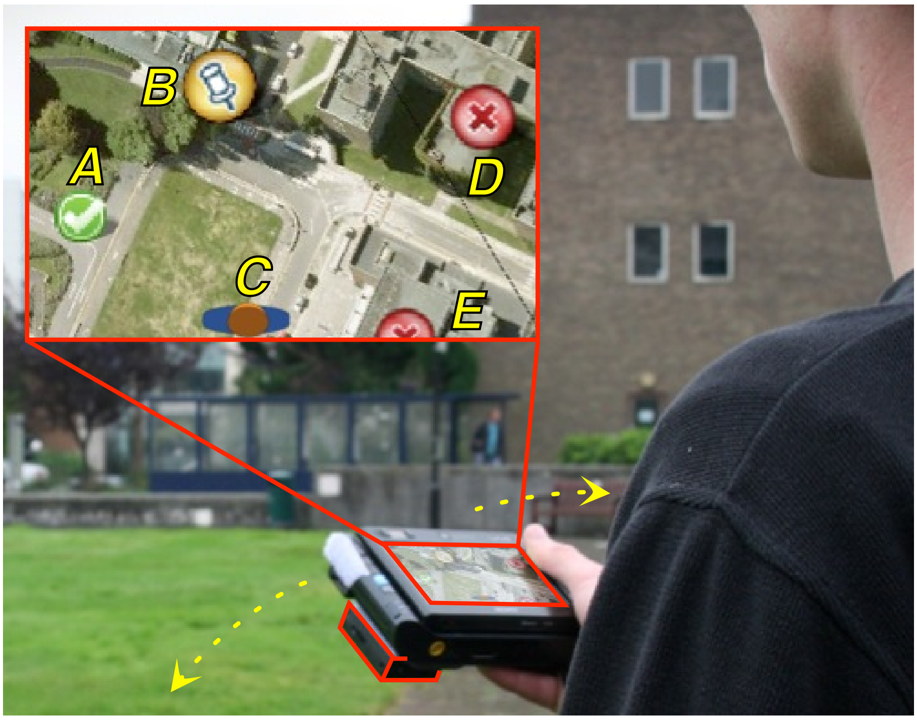


Figure 4.4: The visual browsing comparison system, showing the on-screen display (inset). Point *A* has previously been selected, point *B* is currently being pointed at by the user and is activated. Point *C* represents the user at their current position. The two remaining points, *D* and *E*, have not yet been visited.

haptic system in its implementation, to the extent that it was possible. The system uses the same hardware as the Sweep-Shake prototype, but differs in that instead of offering vibrotactile feedback, the user is given a visual display of their actions.

Similar to the initial browsing functionality of our haptic design, the visual system provides a discovering mode, presenting the user with an aerial photo of their location with markers overlaid to show potential objects of interest in the same field of view (see Fig. 4.4). As in the haptic system the user is able to determine the quantity of information at each location, illustrated by the size of the marker (a larger radius indicates more information is present), and the centre point of the object, shown as a cross at the centre of each icon. As the user turns to face potential targets the map display is re-oriented to ensure that the visible display represents the real-world surroundings as seen by the user.

When a potential target is centred in the user's field of view, it is visually highlighted to show that it is available to be explored. The UMPC has several control buttons positioned around its screen, and we repurpose one of these to allow the user to control the discovering and displaying interactions in an equivalent way to that used in the Sweep-Shake system. Pressing the zoom control button on the UMPC at this stage will switch to the displaying mode, showing a visual zooming in effect. This effect lasts the same length of time as the pulses given in our haptic system, helping to ensure we are able to compare the systems fairly. In a similar way to the Sweep-Shake system, any available information is clustered into type categories and represented by four distinct media icons, each in separate corners of the display. Users can select an information category by touching its icon on the screen of the device.

When the user has finished browsing the information, pressing the zoom button again will return to displaying the cluster of data types, and pressing a second time will return to the initial browsing mode.

Participants

Thirty-two participants aged from 18 to 65 were recruited for a 30 min study. Fourteen participants were university staff members, 18 were students; 16 participants were male and 16 female. Nine participants had previous experience of accelerometer-based interaction methods using a games console; the remainder reported no prior experience of this style of interaction. None of the participants had taken part in the first study of the Sweep-Shake prototype.

Participants were equally split between two study conditions: the Sweep-Shake system and the visual comparison system. Randomly, four of the nine participants with prior accelerometer experience were allocated to the Sweep-Shake system and the remaining five to the visual system. The 23 participants without prior experience were randomly allocated between designs, giving a total of 16 participants per system.

Targets

Before the study began, two sets of six distinct pre-set targets were created within a 100 m radius of where participants would be standing during the study (see Fig. 4.5). Half of these targets were used for an initial training session, the remainder were used for the tasks participants were asked to complete. Each of these second set of targets was also randomly allocated between one and four content categories, with 15 created in total over all six targets. These sets of targets were identical for each participant over the entire study.

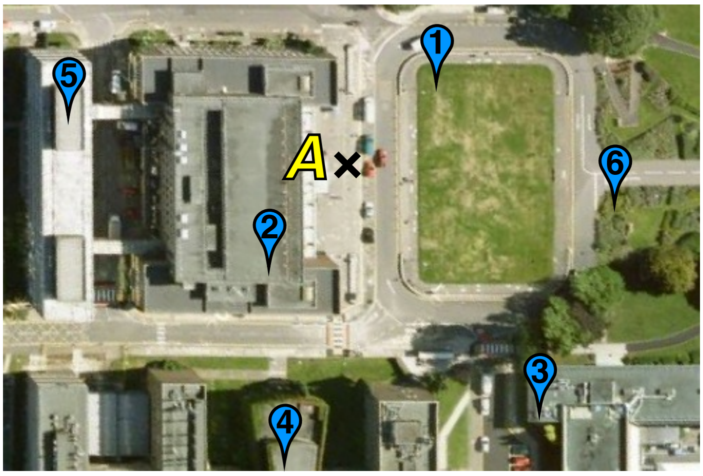


FIGURE 4.5: Targets used for the study. The participant, standing at point *A*, finds, points toward and attempts to display content categories for each target. All targets are visible from the point the user is standing at. Participants found and selected targets in any order they wished – numbers identify targets for analysis only.

Tasks

Participants completed the following tasks during the study:

1. Find and select each of the six targets in the area around them (see Fig. 4.5). When a target was found, participants attempted to:
   (a) Zoom in to display the content categories of the target.
   (b) Find and select each available content category.
2. Complete a questionnaire to rate their usage of the system.
3. Give any verbal feedback resulting from their usage of the system.

Measures

Several measurements were automatically recorded by the system whilst participants browsed and discovered the six targets around them. We recorded the number of targets found and the time taken to find each target, in order to measure the effectiveness of the Sweep-Shake system against the more traditional visual-based design. For the first target the time taken was recorded from the system start time; for subsequent targets this was measured from the when the participant zoomed out from the previous target.

We also recorded the delay between the participant first discovering (i.e., pointing toward) a target and their subsequent selection of the target, aiming to highlight issues with participants' ability to respond to visual or haptic feedback cues. Related to this measure, the number of target activations without selections (i.e., when a user pointed at a target but failed to press the button to select it) was recorded, giving a measure of the system's success in communicating to the user the presence of a target and allowing them to select it. Finally, the system recorded the number of false positives, defined as a participant pressing the button to select a target when they were not actually pointing toward any of the targets.

When zoomed in to display content categories for an individual target, different automatic measurements were taken, focusing on the success of the hierarchical haptic feedback in displaying to the user the different types of content. These measures were:

D1 The number of content categories successfully found and selected from each target, allowing us to measure the effectiveness of the feedback given at displaying each of the four content types.

D2 The time taken to find each of up to four content categories for each target, showing whether participants would be able to find information categories within a reasonable amount of time. For the first content category this was recorded from when the participant zoomed in; for subsequent content categories this was from the time the previous content category was selected.

D3 The number of times participants needed to zoom in to the displaying mode of each target in order to select all of its content categories, which helps to highlight any issues of mode confusion, or areas where repeated interaction is necessary.

D4 The time spent pointing to the haptic area for each content category, giving an indication of whether the feedback given was clear and quickly interpretable *(not applicable for the visual comparison system)*.

D5 The number of times the participant touched or pointed to each content category, indicating whether participants were able to get an overview of the information available about each target without the need for multiple interactions.

Participants also completed a questionnaire based on the six factors of the NASA Task Load instrument (TLX) [56], examining their perception of the costs involved in using the system. They were asked to rate the mental, physical and temporal demand imposed, their success in performing the selecting task, the overall effort needed and their frustration with the system. Each of these dimensions was rated on a scale of 1 (negative; e.g., high frustration, low performance) to 7 (positive; e.g., low mental demand, high performance). In addition, each participant was asked to rate, on the same scale of 1–7, specific aspects of the prototype's usage and usability. The features rated were: their overall ability to identify the actual targets they were marking; how fast they felt they were able to mark the targets; and, the perceived usefulness of the system for finding directions, points of interest, urgent information or for simply filling time.

Procedure

At the start of each study session, each participant was met individually and given an overview of the study and its purpose, followed by a discussion of a short usage scenario. Participants who had been allocated to the visual system were then given a short usage demonstration and, as a form of training, practised both target discovering and displaying by finding and displaying content categories for up to three training targets from a window in our laboratory. Participants allocated to the Sweep-Shake system were first shown a demonstration using the visual system to help illustrate to them the interaction methods. These participants were then given a demonstration of the Sweep-Shake system and were asked to use it to perform the same training exercise as participants using the visual system.

Following this short training session, each participant was then taken outside to a fixed location in an open space on our university campus. Every participant in the study stood in the same location. As an additional training exercise participants used the system they had been allocated to locate and select each of the first set of six targets. The displaying interaction was not used during this training exercise; our aim was to familiarise users with the act of pointing to buildings and other objects to feel haptic feedback related to geolocated content.

Participants were then informed that there were a further six targets to explore, and proceeded to use the system to discover each of the second set of six targets, and then display each available content category. While participants completed this task, the researcher observed their behaviours and methods used in finding and selecting targets. After completing this task, participants completed the TLX questionnaire to rate their usage of the system. Finally, participants were asked for verbal feedback resulting from their usage of the system, and this was recorded by the researcher. Participants were then given a bookstore gift voucher as a token of our appreciation.

4.4.2 Findings

All participants attempted to find (and believed they had succeeded in finding) each of the six targets available. Each participant also completed the rating questionnaire and offered several verbal comments about the system they had been asked to use. We analysed the measures recorded using GLM ANOVA between targets and between systems, for both discovering targets and displaying content categories.

Discovering targets

Table 4.1 shows the mean and standard deviation of each measurement for the discovering modes of each system. Six participants using the Sweep-Shake system found all targets, with the remaining ten participants missing between one and four of the six targets. Thirteen participants using the visual system found all targets, and the remaining three found all except one. There were 177 false positives recorded over all participants: 91 when using the Sweep-Shake system and 86 when using the visual system.

Figures 4.6 and 4.7 show the spread of times taken to find each target, and the activations without selections that occurred. Participants using the Sweep-Shake system often activated targets without selecting them, instead scanning the area and returning later to select targets. This behaviour increased the time taken by

Measurement	Sweep-Shake		Visual	
Number of targets found (of 6 available)	4.5	(1.5)	5.8	(0.4)
Time to select each target after discovery (s)	16.5	(22.3)	8.8	(5.6)
Time between target activation and selection (s)	1.7	(1.4)	1.9	(1.6)
Activations without selections (per target)	9.3	(6.0)	0.7	(0.5)
False positives (per target)	0.9	(1.1)	0.9	(0.6)
Total time taken (s)	105.2	(32.3)	81.7	(26.4)

Table 4.1: Means and standard deviations (in parentheses) of each of the measures recorded when browsing, and the total time taken to complete the task.

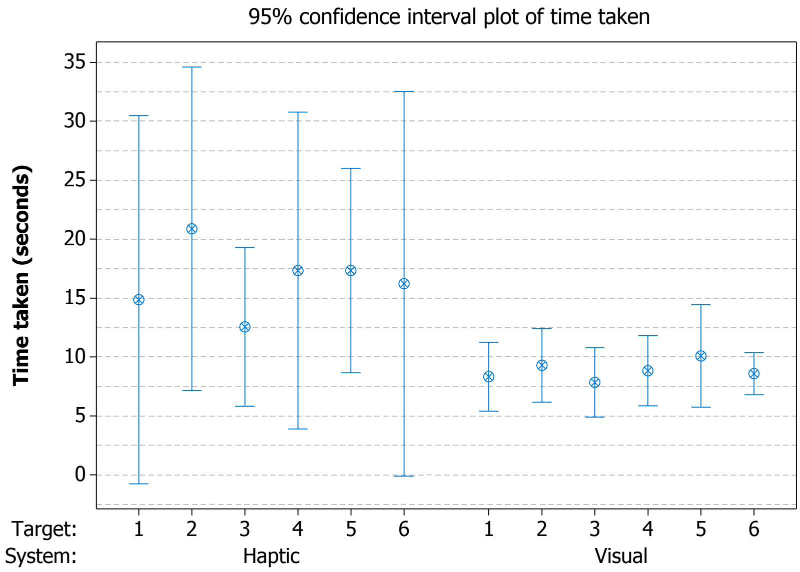


Figure 4.6: The time taken to discover each target for both systems.

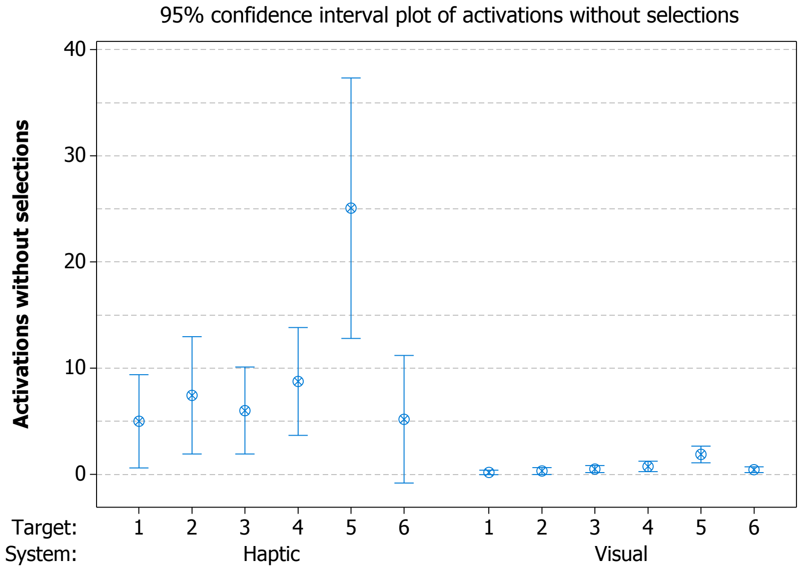


Figure 4.7: The number of times targets were pointed to (activated) without the participant actually selecting the target.

participants using this system. No significant difference in times taken was found between targets when using the Sweep-Shake or visual systems individually ($p > 0.05$), but when comparing systems there is a significant time difference between the Sweep-Shake and visual versions ($F = 10.32, p = 0.002$): participants using the visual system took significantly less time to discover targets. For target activations between systems there is a significant difference ($F = 37.67, p < 0.001$): the visual system caused significantly fewer activations without selections.

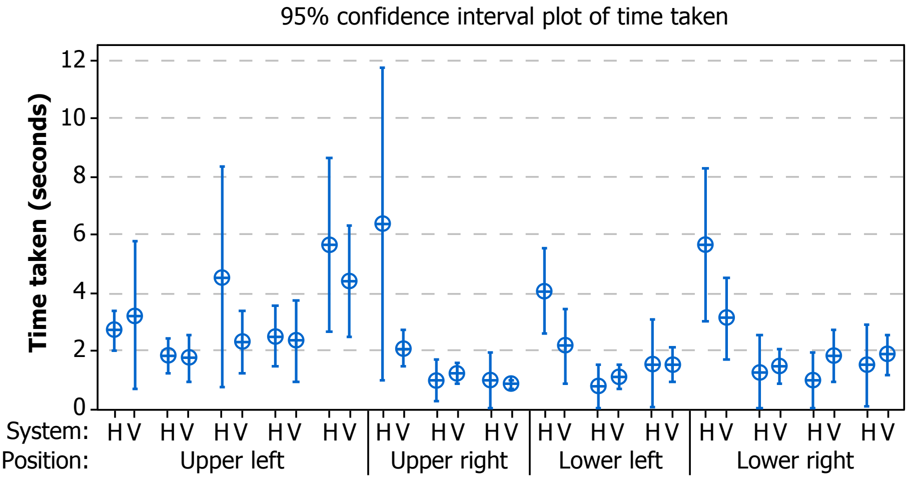


FIGURE 4.8: Times taken to find each of the fifteen available content categories over all six targets using each system, grouped by the position in which they appeared to participants.

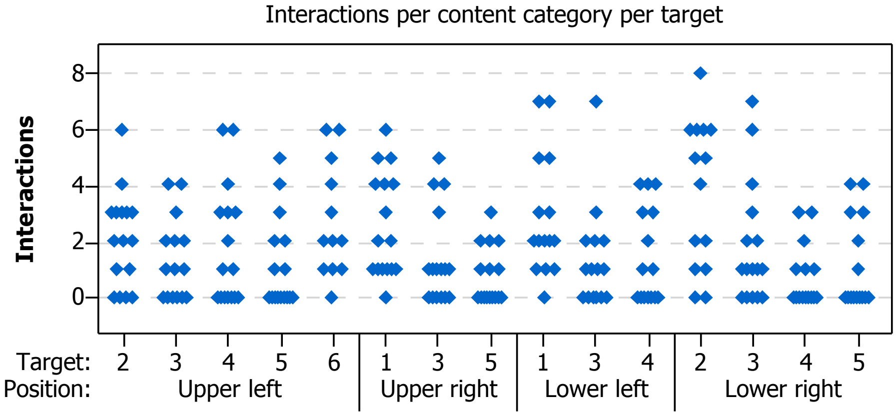


FIGURE 4.9: The number of times each participant pointed to each of the content category regions for each target while using the Sweep-Shake system. Participants often pointed several times to a content category before ultimately selecting it.

Measurement	Sweep-Shake		Visual	
D1: Content categories selected (% of 1–4 per target)	72.6	(43.5)	99.0	(5.4)
D2: Time taken to find each content category (s)	3.0	(4.5)	2.1	(2.3)
D3: Zoom interactions required (per target)	1.0	(0.2)	1.0	(0.4)
D4: Interaction time with each content category (s)	8.4	(11.1)	*n/a*	*n/a*
D5: Number of interactions (per content category)	3.5	(3.7)	1.0	(0.2)

TABLE 4.2: Means and standard deviations (in parentheses) of each of the measures recorded when displaying content types.

Displaying content categories

Table 4.2 shows the mean and standard deviation of each measurement for the displaying modes of each system. Participants using the Sweep-Shake system found around 73 % of the available content categories on average, while participants using the visual system found almost all of the content categories.

Figure 4.8 illustrates the time taken to discover each content category over both systems. Content categories are grouped into the four positions in which they appeared to the user: upper left (text); upper right (images); lower left (video); and, lower right (audio). In many cases the difference between systems is minimal, but for some there is a larger variance for the Sweep-Shake system.

When comparing individual results between systems there is a significant difference between systems for the first lower left region ($F = 4.36, p = 0.046$): the haptic system took significantly longer to find this content category for this target. The visual differences between other content categories are not significantly different at the 95 % confidence level ($p > 0.05$). There is no significant difference between systems if we compare within the four regions at which content categories appeared to participants, but when comparing the systems overall, there is a significant difference ($F = 8.21, p = 0.004$): the visual system was significantly faster overall for participants to select content categories.

Figure 4.9 shows the number of times each participant pointed to each content category when using the Sweep-Shake system. As in our study in Chapter 3 of feedback-assisted discovery of targets, participants have pointed to each tactile region several times before selecting. Interestingly, some participants pointed to category regions (i.e., triggered feedback), but did not subsequently select the category.

Subjective ratings

Figure 4.10 shows the average TLX ratings given by participants. Significant differences were found between the two systems when considering performance ($F = 11.67, p = 0.002$) and frustration ($F = 6.02, p = 0.02$): the visual system was rated significantly better on these aspects. Figure 4.10 also shows the ratings given for system usage and usefulness. Significant differences are evident for identification accuracy ($F = 19.35, p < 0.001$) and speed of marking ($F = 12.45, p = 0.001$), with participants rating the Sweep-Shake system lower on these aspects. No further significant differences were found when considering any of the remaining ratings.

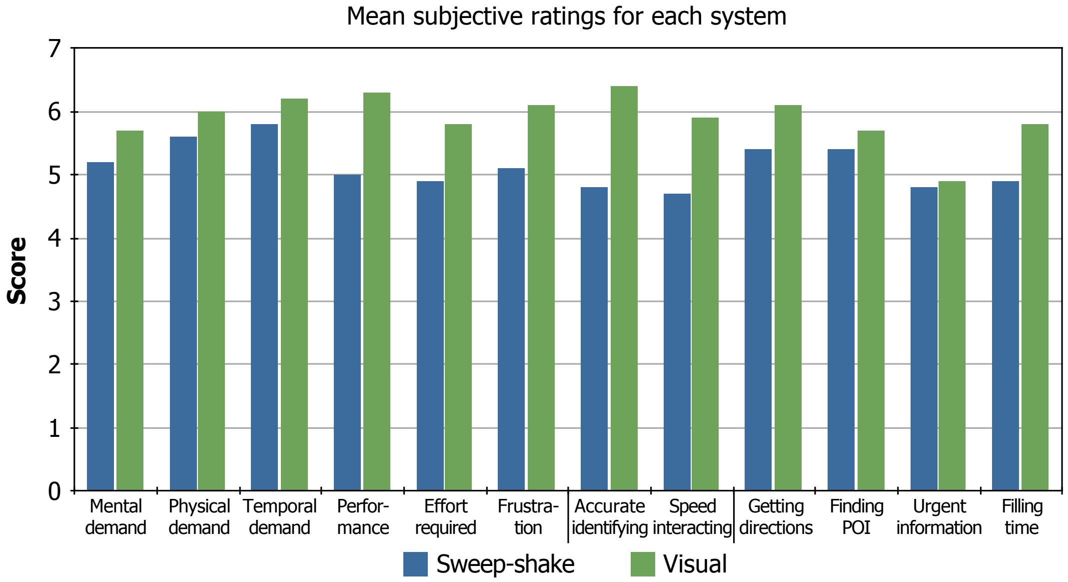


FIGURE 4.10: Subjective system usage ratings. Left: participants' scores for each of the six TLX questions. Centre and right: ratings for system usage and usefulness of the system for specific tasks.

Verbal feedback

Four participants of the 16 using the Sweep-Shake system said they would like to be able to use the device in their everyday lives, and two of these thought that using the system would get much easier with practice. Three participants specifically stated that using the Sweep-Shake system was fun, and five others said that the act of pointing at a location was very easy and they could get used to it quickly. One user commented that the system was "much more helpful than my GPS for finding places," but that it would be harder when the information space was cluttered, a statement echoed by three others.

Four participants found the Sweep-Shake system's different vibration pulses hard to distinguish from each other, and said that they had not been able to detect the increase in tactile intensity at the centre of each target. A further three participants reported difficulty in remembering or tracking which mode the system was in when attempting to display content types. Six participants said they appreciated the general idea of the device, and specifically mentioned the same 'guide me' mode suggested by participants in our earlier study, but one participant was concerned about the social acceptability of pointing to objects with their mobile phone.

Four participants of the 16 using the visual map system commented that they found it easy to use, and that they liked being able to see the buildings around them from above. One participant stated that they found the discovery of each target easy primarily because they were familiar with the test area – had the study been conducted in a new location then they felt they would have had less success in completing the task. Three participants using the visual system found it to be fiddly and awkward, and struggled to use the re-orienting display. In addition, nine participants said that they had experienced problems seeing the display of the UMPC due to sunlight reflecting on its screen.

4.4.3 Discussion

Participants in this second study of the Sweep-Shake system have been able to find targets and display content categories with only vibrotactile feedback to aid them. The ability of participants to explore the available content categories via haptic feedback suggests potential for more successful usage of this type of system after further user training and exposure. Indeed, as before, it is also important to consider these results in light of the low degree of familiarity with haptic interfaces. The Sweep-Shake system uses a novel, unfamiliar interaction style, and offers much lower resolution feedback than the visual interface, but participants have still been able to use it to successfully discover and display content from most targets.

Participants using the visual system were able to discover and display more targets, and in a faster time than those using the Sweep-Shake system (8.8 s versus 16.5 s). This difference in speed and accuracy, although clearly a negative aspect for important, time critical information access, can also be seen as a positive point for the types of interaction we are aiming to support. As we saw in the first study of this system, participants are stimulated to explore the tactile areas in their environment whilst simultaneously taking in the visual scene in front of them, rather than concentrating on the digital representation on their mobile screen. This stimulated exploration is perhaps illustrated in this study by the number of activations without selections in our results – participants seem to have skimmed over the available targets before deciding to go back and explore them in more detail later. Similar to our vibrotactile system in the previous chapter, participants using the Sweep-Shake system seem to have explored slowly, browsing targets one-by-one in a more reflective approach than those using the visual alternative. It seems likely that this type of interaction would match closely with the delayed searching approaches that Brown et al. [17] and Jones et al. [73] motivate, while a visual display is often more appropriate for realtime content display.

Participants had difficulty in selecting some targets when using both the Sweep-Shake and visual systems – the feedback for target five in particular was activated many times without the target being selected. One possible explanation for this is its distance from the user (see Fig. 4.5). As we saw in Chapter 3, targets further away from the user take up less of their field of view so require a more accurate pointing gesture to select. We did not expect this issue at the shorter distances used for this study but it suggests a need for our systems to compensate for distance when determining the size of the initial vibrotactile feedback area.

The participants who used the Sweep-Shake system seem to have had difficulty identifying the mode they were in at times, and false positives and verbal comments to this effect resulted. Clearly this is an area where further haptic feedback development could offer usability improvements. While it is relatively easy for users to assess mode changes in visual systems, a richer set of haptic forms may be needed to clearly communicate state shifts. Surprisingly, this issue was not raised in our first study of the system, suggesting that the presence of real location data helps to alleviate difficulties in determining the system state. It is likely that removing the actual information made the task more difficult for participants using the Sweep-Shake system.

When asked to rate the systems, the visual display was scored only marginally higher than the haptic system. In their verbal feedback, too, participants seem to rate their usage of the Sweep-Shake system similar to that of the visual system, with several comments about ease of use and the natural feel of pointing to locations. It was interesting that several participants mentioned the use of tactile feedback applied to a familiar GPS-assisted navigation task. This suggestion, which was also mentioned by two participants in the initial study of this system, indicates that, for pedestrians, it might be appreciated if wayfinding was more of a background task, rather than a prescriptive turn-by-turn instruction stream as in current navigation devices.

Previous work (e.g., [76, 141]) has investigated the use of audio feedback to help indicate when a pedestrian leaves a pre-defined path. Visual or audio-based versions of these types of systems are commonly used in tracking beacons, though possible issues can arise when obstacles have to be surmounted and routes are indirect (see [76]). Comments from participants in this study suggest navigation might be better when given on request, rather than via the common implementation of pushing all available information whenever it could be useful.

Participants described this interaction as a 'guide me' mode, and this is an obvious area where haptic feedback could offer a more eyes-off, physically-grounded interaction style. From the research discussed so far in this thesis we have seen that tactile feedback can offer a relatively precise directional indication, but without the need to look at a screen or wear headphones. A tactile navigation system, then, seems ideal for helping users request the direction of a location beacon rather than follow turn-by-turn guidance.

There are two approaches to this type of navigation, however. The first option is to consider the feedback that is given as a method of displaying the location of some end point that a person is aiming to reach, but leaving actual navigational choices entirely down to the user. The second option is to focus more on the navigational behaviour of the user, and attempt to shape the way they interact with and move through the space they are in. In the rest of this chapter we concentrate on the first of these two approaches. Rather than displaying a static navigational end point, however, we attempt to help a group of pedestrians find a mutually-convenient dynamic meeting point. The second of these approaches is explored in Chapter 5, where we turn to investigate how the movements of people through a space can be shaped by the navigational device they use.

4.5 Vibrotactile display of dynamic elements

From our studies of the Sweep-Shake system it is clear that people are able to use vibrotactile feedback for displaying static content. However, feedback from participants in these studies suggests that navigation, rather than content display, might be a more attractive prospect, in some cases. As we discussed in the introduction to this chapter, representing the locations of people via, for instance, their central point, is an example of displaying geolocated elements that are dynamic, rather than fixed in place. This dynamic central point is an ideal position for participants to aim for when they attempt to meet up.

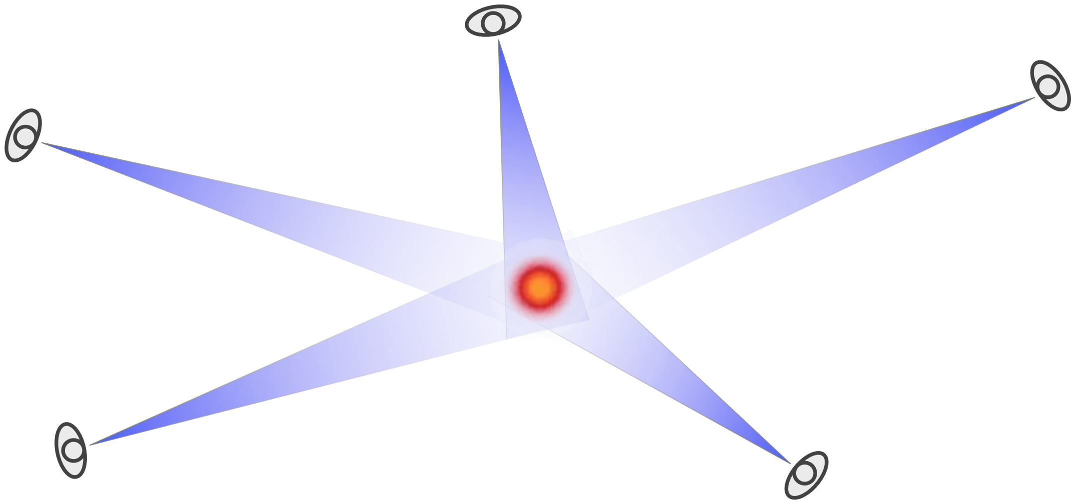


FIGURE 4.11: A group finding their way to a mutually-convenient meeting point using our prototype system. Each participant in the group feels vibrotactile feedback when pointing toward the group's central point. The beacon's position is dynamic, adjusting as the positions of each member change.

Previous work has investigated group behaviours when attempting to meet up, but these have used visual-based systems. Axup et al. [5], for example, used text messages to co-ordinate a rendezvous, but found that the visual attention required to operate the system forced users to alternate their attention between the screen and the environment to avoid walking into obstacles. Colbert [29, 30] showed that rendezvous in unfamiliar locations required more communication than familiar locations, and Dearman et al.'s study of pairs rendezvousing [35], found that the majority chose a meeting location that was a middle point, with only one pair choosing a landmark. Further work by Dearman et al. [36] found that participants often maintained continual awareness of partner and meeting locations throughout the process.

In this section we describe the design and evaluation of a prototype designed to use tactile feedback as an aid for dynamic group rendezvous. We aim to address the issues highlighted in previous work, and also to show how the display of a group's central point can be used as a successful cue for supporting meetups.

4.5.1 Prototype design

Figure 4.11 illustrates our prototype system design. As before, users scan for directional tactile feedback on demand. However, in this prototype, while people scan for feedback their locations are continuously sent wirelessly to a remote server. The server computes the central point of the group, based upon a spring model, and reports this back to each client, which uses this new point as the haptic target.[5] Each person, then, is scanning for, displaying, and then moving towards a dynamic group central location, rather than a fixed position.

Participants in the two previous studies described in this chapter reported that they could not feel the differences in feedback intensity between the centre and edges of each target, but they were still able to find targets; consequently, we removed the variable feedback intensity element of our previous systems. Our design also aims to support typical user behaviours during a meetup task, as reported in previous work – foremost by providing a mutually-convenient central location at which to rendezvous [35]. The eyes-off form of our prototype is intended to combat previously reported issues with visual attention, but still support meeting point awareness when desired [5, 36]. In addition, in our design no group communication is necessary, and although local knowledge may shorten the time taken to meet up, it is not required [29, 30]. Finally, while the system provides meeting point awareness on demand, it does not provide explicit group member locations, in order to preserve each person's privacy during the meetup process.

Simulating group rendezvous

Our design uses the same type of directional tactile feedback as used in our previous designs, but for this prototype the parameters used were refined in a simulated model of both user behaviours and environmental uncertainties. This simulator, developed by our project partners and released in open source form as part of this research,[6] aims to model simple navigational tasks while also incorporating variable external constraints. While a simulated environment cannot capture many of the subtle complexities of human behaviour, the major uncertainties in the system can be reproduced, and the effects on navigation performance observed and quantified.

The model's configurable parameters, specifically: feedback zone width; walking speed; time taken to scan; variation in walking rate; user's heading adjustment rate; network communication delay; GPS noise; and, GPS update rate, are used to guide

---

[5] The development of this prototype was predominantly by our project partners (as detailed in Appendix A, [P6]), hence its design and rationale are not described in great detail here. Further system details are available in [161].

[6] See http://www.negotiatedinteraction.com/socialgravity

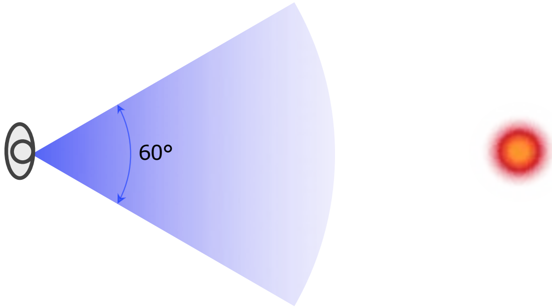


FIGURE 4.12: The 60° haptic zone width used for the prototype (actual size). Feedback is activated when a user points their device within the shaded area. The centre of the zone aims directly at the goal.

simulated users around a model of a real space. The simulator allows reproduction of hundreds of repeated trials—costly and usually entirely infeasible with user studies—and reports overall statistics to help evaluate the success of the parameters chosen.

The prototype's design was refined in this manner, and the simulator showed that participants would likely be able to meet up efficiently even with a poor GPS fix and with network delays exceeding 30 s. Various angular widths of the haptic zone were evaluated, and a value of 60° was eventually chosen as an optimal tradeoff between sensitivity and efficiency. Figure 4.12 illustrates how feedback is activated.

We created an accurate model of the environment in which the system would be evaluated, and used this for 500 simulated runs of the rendezvous task using the exact positions of the starting points of the participants in the field trial. The simulator predicted a rendezvous time of 17 min 45 s (s.d. 5 min 37 s), for five people walking at 0.8 m/s (±0.2 m/s), or 2.88 km/h (±0.72 km/h). The simulation modelled Gaussian GPS noise (s.d. 8 m) and Gaussian angular noise (s.d. 8°), checking for feedback every 30 s.

4.5.2 Experiment: exploring vibrotactile-supported rendezvous

We undertook a field trial to examine potential usage of the system in a realistic scenario. Our research questions were:

Viability Can the system be used by a group of pedestrians to help them meet up?

Impact What is the effect of using the system on the participants' behaviour, compared to walking behaviour without the system's guidance?

Five separate one-hour trials were performed to help understand these questions. Twenty-five participants aged from 18 to 65 were recruited for the trials, in five groups of five people. Fifteen participants were male, ten female; six students and 19 members of university staff. None of the participants worked in areas directly related to HCI.

Tasks

Participants completed two primary tasks during the study:

Rendezvous From an initial starting position, use the device to scan for the location of the meeting point, then attempt to find this, using their own judgement at any path choices until meeting up with other participants.

Free walking Repeat the task, but this time use their own choice of route to the same meeting point, with no feedback to guide them. Completing this second task allowed us to compare between participants' normal behaviour and that when using the system.

It is important to note that participants were told that they should not consider either of the tasks as a 'race.' Participants were asked to imagine that they were walking, chatting to friends or looking at the sights around them, rather than focussing solely on feeling for vibrations and reaching the goal.

Measures

A large quantity of data from both logs and observations were collected from each hour-long trial, allowing us to measure the success of the system against each of our research questions. Each participant was also observed while using the system, and each group of participants was asked about their participation in a semi-structured interview after each trial. Each measure was recorded specifically to help answer one of our research questions.

Viability Measured as the percentage of group meetups that were successful. In addition, we measured the time taken to meet up relative to the time predicted in the simulation, to see whether groups met in a realistic amount of time.

Impact We measured the impact of the system on participants' normal behaviour by comparing the two routes taken to the meeting point, both in time and distance. Examining the time spent scanning for the meeting point compared to the time spent walking without scanning allowed us to assess the impact of the act of scanning for feedback. As in the previous chapter, we also measured the effect that using the system had on each participant's walking speed. In the outdoor study area used for this trial we were unable to compensate for the inaccuracy of pedometer-based walking speed measurement by measuring distance as previously, so compensated for the increased variation in the outdoor environment by using a much higher resolution analysis of sensor data than is commonly used. This allows a comparison of behaviour in both the rendezvous and free walking cases down to the level of individual steps, and at the different stages and conditions throughout the experiment.

Procedure

At the start of each session each participant was met by a researcher and given an overview of the system and its purpose. Each participant's meeting location was separate to minimise effects from participants recognising each other before meeting up. Participants then used the system for a short training session, lasting no more than five minutes, in which they were able to feel example feedback and use the system while moving. Following this, each participant was taken to a starting location at the edge of the university campus. The five starting points were the same for each session, and were spaced evenly at the edges of an area of approximately 0.5 km^2 (see Fig. 4.13).

When all participants were in place, each began the rendezvous task, following the vibrotactile feedback while a researcher observed their behaviour from a short distance behind them. The researchers did not interact with participants during the tasks. When all participants met up they were led back to their starting points and asked to make their way to the rendezvous point a second time, this time using their own choice of route rather than using the feedback to guide them. This free walking task provides a baseline measure of the best possible performance where users know exactly where to go and do not need to interact with the system.

Measure	Rendezvous	Free walking
Time taken (min:s)	13:05 (1:50)	7:44 (0:55)
Distance walked (m)	992 (193)	573 (161)
Walking speed (m/s)	1.24 (0.34)	1.28 (0.60)
Walking speed (km/h)	4.46 (1.22)	4.61 (2.16)

Table 4.3: Summary of mean results for the rendezvous and free walking tasks (standard deviations in parentheses). The difference between the two tasks is almost entirely due to the larger distance walked in the rendezvous case.

Finally, when all participants met up for the second time a short semi-structured interview was conducted. At the end of the study each participant was rewarded with a bookstore gift voucher as a token of our appreciation.

4.5.3 Findings

All participants successfully completed the feedback-guided rendezvous task. They were also all able to make their way back to the meeting point after being returned to their respective starting points. Table 4.3 shows the results for the time taken to meet up, and also the distance walked and walking speeds for both tasks.

Participants took around 13 min to rendezvous while using the system to guide them. The time taken for participants to return to the same point a second time was lower, but this time difference is almost entirely due to walking further in the rendezvous case. Participants' walking speeds were not significantly different between the two tasks. The extra distance between tasks may be due to the limited ability of participants to plan efficient routes around obstacles in the rendezvous case. However, it is important to remember that in the free walking case participants were navigating from the starting point to a known destination, and could take advantage of any knowledge of the environment or shorter routes.

In both the rendezvous and free walking cases participants' walking speeds were faster than the 0.8 m/s (2.88 km/h) we chose for simulations. This is the main cause of the difference between simulated and actual times taken – scaling the 13 min rendezvous duration from the trial to 0.8 m/s gives an approximate duration of 20 min, which is far closer to the simulated result of around 18 min.

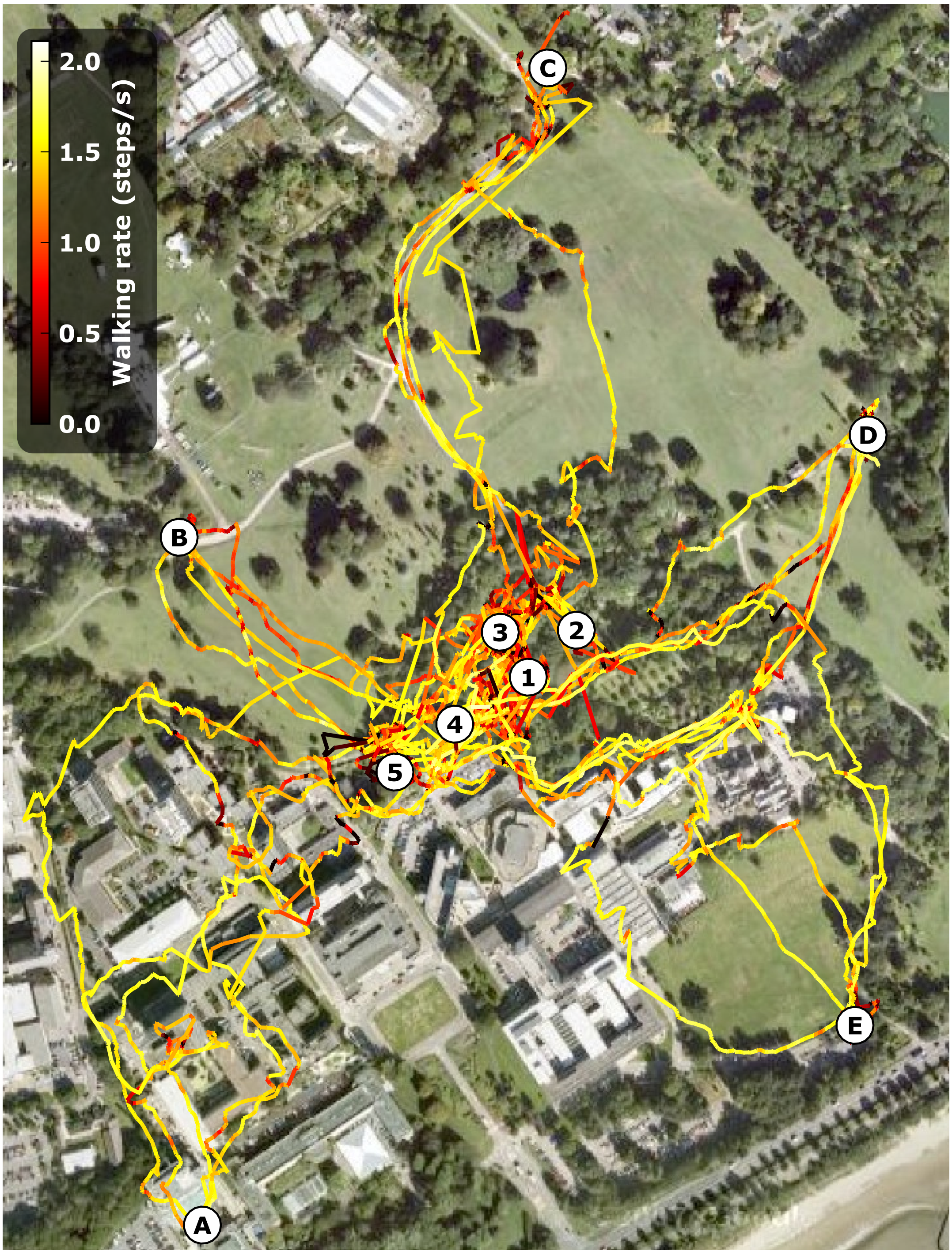


FIGURE 4.13: Paths taken by all 25 participants during the rendezvous task, starting at A–E, ending at 1–5. Lines are coloured according to the participant's walking speed for each part of the route (steps/s, see key). Generally people walked at a constant pace, with occasional stops. Stopping was more frequent in the vicinity of the rendezvous point.

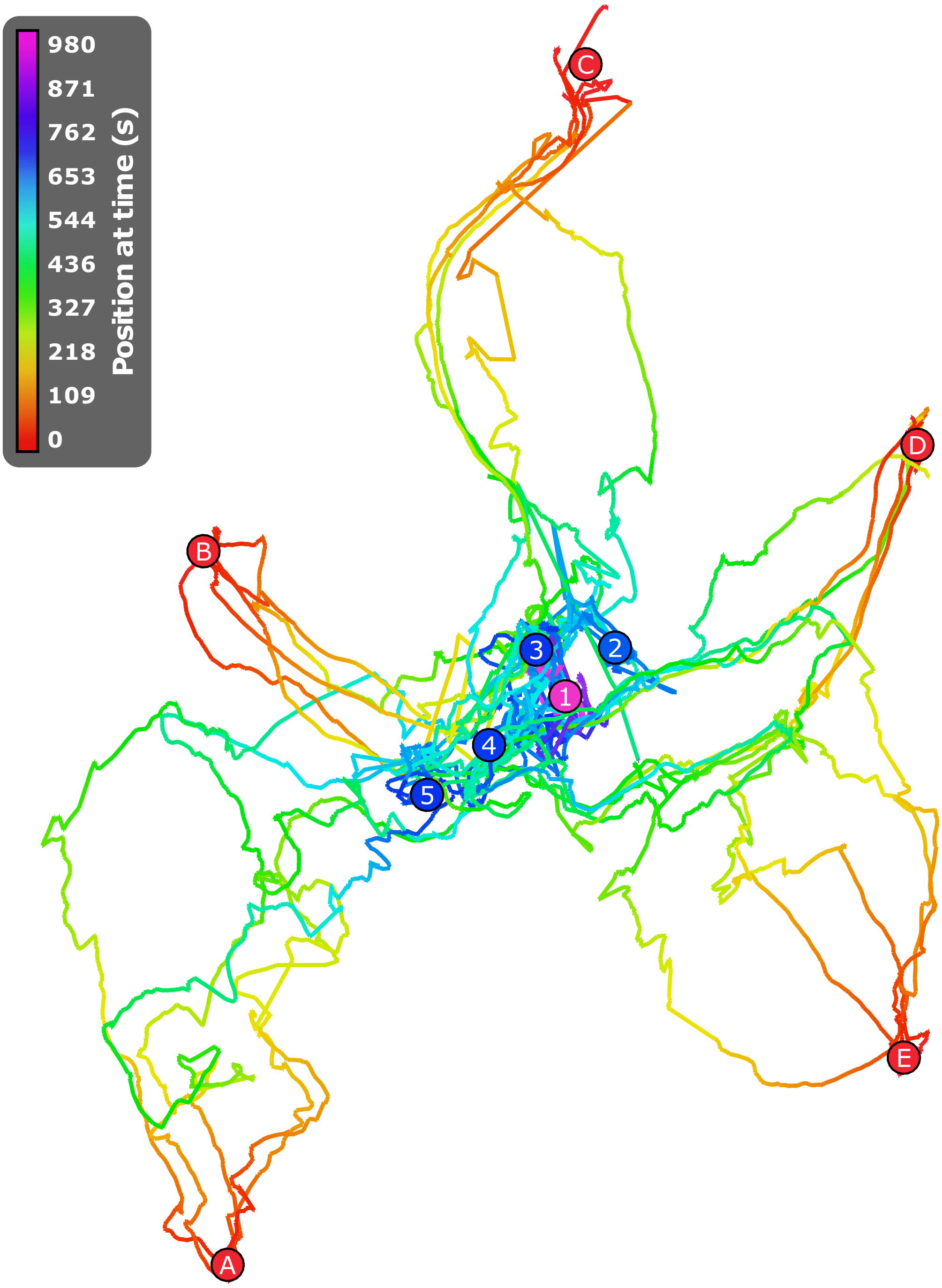


FIGURE 4.14: Positions of all 25 participants throughout the rendezvous task, starting at A–E, ending at 1–5. Lines and points are coloured according to the participant's location at any given time (s, see key) during the task. Generally, as also indicated by Fig. 4.13, although participants took different routes, they progressed at a similar rate toward the goal.

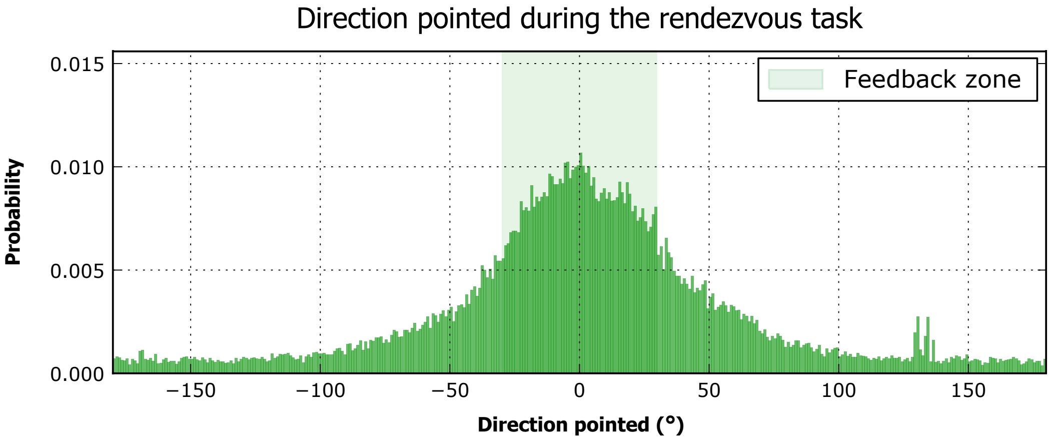


FIGURE 4.15: The directions participants pointed throughout the study, relative to the direction of the zone in which feedback could be felt. Participants tended to feel in or around the feedback zone most of the time.

Walking behaviours

Figures 4.13 and 4.14 illustrate the paths chosen by and positions of participants for each rendezvous trial, coloured according to their walking speed and location along the route. Walking speeds are similar for a large part of each route, with most participants slowing down only when near the eventual meeting point.[7] Looking more closely at the individual routes taken by participants, we can see bottleneck areas where location constraints have limited path options, forcing participants along similar routes. When route options are more open, the paths taken are more varied.

In some cases participants stopped or reversed their steps at various points during the route. This could be for a variety of reasons: encountering a dead end or obstacles, a change in the position of the central point, or perhaps just a desire to take a different route. Regardless of the reason, we can see that participants have been willing to take diverse paths. This suggests that they trusted the feedback sufficiently that they were able to take alternative routes, knowing that they would be brought back on target.

Figure 4.15 illustrates the directions in which participants pointed throughout the task, relative to the feedback zone. Participants preferred to keep their focus on the vibration zone for most of the task. The cluster of readings around 140° was caused by a short period of Bluetooth connection problems between devices; consequently the system received unreliable heading readings for approximately 10 s. There is no

---

[7] A comparison of participants' walking behaviours to the system simulations is available in [161]. Here we focus on participants' walking behaviour and their ability to use the haptic display of a central meeting point to rendezvous.

way to accurately tell where participants were actually pointing during this period, but we have opted to show this discrepancy rather than smoothing the visualisation.

Participant observations

All except two participants held the device in their hand by their waist, while two chose to scan with their arm held almost horizontal. Most participants scanned for feedback constantly. Several participants encountered locations where they were unable to directly follow the feedback due to an obstacle in their way; in this situation they continued to scan for feedback to one side for reassurance.

In three of the five trials up to three of the participants met each other before the group rendezvous. In two of these smaller groups participants struck up a conversation and appeared to be scanning for feedback much more casually than they had been while on their own, only stopping to explicitly scan when they reached a point where several potential paths were available, and often not stopping at all. The third group of participants walked together, but without talking to each other, appearing to still direct their attention toward finding the direction of the feedback. In one trial several participants met very near to the eventual rendezvous point but, interestingly, rather than continue together, chose to follow the feedback from their individual devices. Due to GPS inaccuracies the feedback given was in slightly contradicting directions, so the participants separated again and walked down different paths.

Post-study interview

The interviews recorded after completion of the rendezvous tasks provided valuable insights, and many suggestions for potential future use of the system, including: festivals and large gatherings of people; when in unfamiliar locations; solo navigation to a fixed location; an adaption for blind users; as part of a large area game; in lieu of calling (to avoid high costs). The majority of the participants said they would definitely use the system if it could be incorporated into their usual mobile phone, but several pointed out issues with the final part of the rendezvous task. When very near to each other, participants had often been given conflicting directions due to GPS uncertainty. A commonly-suggested solution was a form of alternative notification to alert users that they were near the final meeting point.

Four participants were completely unfamiliar with the study area but had no problems using the feedback to navigate to the rendezvous point. Four participants noted issues in trusting the system – over time they felt they could begin to gain trust, but in a short study they were cautious about accepting the directions given. Two participants said they had felt they had to adapt their behaviour to cope with the slight lag in vibration response, finding that they sometimes needed to scan several times to get an accurate indication of the direction they needed to head toward. This finding, similar to that found for our haptic system in Chapter 3, highlights the importance of minimising feedback delays in this type of system.

4.5.4 Discussion

All participants in this study were able to use the system to meet up, with only directional vibrotactile feedback to guide them. The system helped participants rendezvous efficiently, and without needing to use a screen-based interface. This eyes-off approach allowed them to navigate without the need to look at the device; consequently participants were able to walk while looking at their surroundings or talking with other people. The use of dynamic feedback allowed participants to locate a mutually-convenient group central point without the need for prior organisation or discussion, and participants were able to track this moving tactile zone effectively, using it to help choose appropriate routes.

Developing a simulator, using it to test design scenarios, and then evaluating the real system using these scenarios proved to be an effective technique in minimising the potentially fiddly and time-consuming configuring of physical prototypes. The low resolution of the vibrotactile feedback that was chosen from the simulations did not greatly affect participants' behaviour. Participants used the tactile cues to meet up, but they did so intelligently, without getting stuck behind obstacles. The use of minimal feedback seems to have been successful in avoiding undesirable situations where users place excessive trust in guidance information, and follow a 'blinkered' path to the goal, ignoring their common sense. Instead, in our design, participants were prompted to choose appropriate routes based on both the system's guidance and external factors that the system was unaware of.

Participants interacting with the large feedback zone used in this prototype did not report the frustration that is perhaps evident in our initial directional feedback experiment (see Chapter 3, Section 3.3), where some participants scanned around and probed the tactile zone repeatedly before selecting. The large feedback zone also shows surprisingly minimal effects on user performance – participants have used their initiative to determine their route from the feedback given, without requiring a precise directional instruction. As might be predicted, the system was of less use for the final stage of navigation when participants were close to each other, mainly due to low GPS precision, which caused inaccurate feedback. Interestingly, in one trial this provoked participants to continue to follow the device's feedback despite having already met up with other users. Refinements are needed for this final stage of the navigation process.

Previous research [36] suggested that users appreciate a constant update of positional information, and this was reflected in our results, with most participants preferring to constantly feel the feedback as they moved. In the study, it is probable that this behaviour was influenced by the specific request to use the device to meet up. Had we given participants a distracter exercise then scanning may have become a background task, as participants in our earlier study suggested. In addition, users' lack of familiarity with the system, and with haptic feedback systems in general, no doubt affected their behaviour. If participants had been more experienced it is likely that the constant monitoring behaviour would decline as their confidence increased. In our own informal pilot studies early in the development of this system, experienced users successfully met up while scanning on a far more intermittent basis (every 90 s approximately).

4.6 Conclusions

We began this chapter by investigating the potential for personal projection to help support physically-grounded displaying. We then turned to investigate how a vibrotactile hierarchy might offer users a broad indication of the types of content available in a location. In the final part of this chapter we returned to investigate a single level of tactile feedback, but extended this to allow users to display the central point of a group of users – a dynamic form of tactile displaying.

Our first prototype in this chapter allowed people to project content discovered via tactile browsing, but there were issues with brightness in realistic settings. Future pico projectors will need a significant increase in brightness and battery life in order to be usable in typical mobile scenarios; the social implications that pico projection in public places will bring are yet to be fully determined.

Projection was replaced, in our novel second prototype, by tactile content categories. Our two studies of this system showed how participants were able to scan as a background task, then display, select and interpret content types using vibrotactile feedback. The results show that the feedback hierarchy clearly allowed people to find and filter virtual targets in their physical environment, eyes-off, though the system was not as effective as a visual representation of the same scene. Further work is needed to improve and clarify the tactile cues given, particularly so that users can sense the change in level from discovering to displaying. From our results, it seems likely that it might be more appropriate to use this eyes-off tactile hierarchy as a slower, more reflective scanning and browsing experience, similar to the delayed searching of Jones et al. [73]. Alternatively, haptic feedback could be combined with other techniques, such as audio, or minimal visual cues. Previous investigation has demonstrated audio-haptic integration – Chang and O'Sullivan [22] found that basic couplings between sound and haptics enhanced the user experience, while Ahmaniemi et al. [1] evaluated dynamic audiotactile feedback for gestural interaction, finding that the combination of haptic and audio feedback improved the accuracy achieved by participants when attempting to perceive the difference between vibrotactile feedback textures.

Our third design in this chapter demonstrated a novel dynamic mobile meetup system that helps groups of users rendezvous in a mutually-convenient location. We demonstrated that only very crude angular feedback is needed for efficient target finding in dynamic pedestrian navigation, even in constrained environments. The use of simulations for studying likely behaviours in mobile trials has offered clear benefits throughout the design and evaluation of this third prototype. A simple simulation of the system in its target environment made useful predictions about behaviour at little cost, and allowed effective design of both the system and its evaluation.

Groups of five participants in our study successfully navigated through a complex environment while negotiating dead ends and moving meeting point positions. Users were able to interact, eyes-off, while walking at a comfortable speed, and the need to navigate and choose routes did not significantly disrupt their normal movement. The process of agreeing a mutually convenient meeting place via SMS or voice channels would take many minutes for this number of people, and would be very difficult in unfamiliar or featureless environments. In our study, groups were able to rendezvous in only three minutes more than it took to walk to a *known* point.

4.6.1 Designing for eyes-off, physically-grounded displaying

The experiments described in this chapter have helped to show how tactile feedback can support complex interactions via straightforward pointing gestures. A vibrotactile hierarchy can allow broad eyes-off displaying of content types, but its accuracy is dependent on users being able to reliably detect mode changes and recognise the meaning of particular vibration patterns. Accurately targeting and interpreting several elements within a short period of time, as in our study, may be difficult in practice. Future systems that combine the haptic hierarchy technique with audio or minimal screen-based feedback might achieve greater accuracy and comprehension, while still preserving the eyes-off nature of the design.

Dynamic tactile feedback, with users tracking a moving target via directional vibration, can support efficient target finding and navigation to a group meeting point. Users are able to interact with a large feedback zone with minimal effect on their normal walking speed. The process of manually agreeing a meeting point is currently time consuming for large groups – while it is likely that our naïve centre point technique will be less effective over very large areas, for situations where time, mutual convenience or privacy preservation are important, dynamic tactile group navigation offers many benefits.

Hierarchies Tactile feedback can give a broad indication of content types via a two-level hierarchy of vibrations. However, careful feedback designs are necessary to ensure that mode confusion does not become a problem – with more than two levels of feedback this is likely to become a more significant issue.

A broad eyes-off display of content types is potentially less attractive to users for categories about which a quick overview is difficult (such as video and audio). Tactile hierarchies might be better focused on delayed interaction, where selection and browsing occur at different times, and so eyes-off immersion in the physical world can be maintained.

Feedback The size of feedback zones, and the delivery of feedback within these, have a large impact on the user experience. A feedback area around 60° in width offers the best tradeoff between over-sensitivity and efficiency in a navigation task. Smaller areas, as used in our earlier prototypes, can be more difficult for users to target. A fast response of tactile feedback might allow the tactile zone's size to be reduced while maintaining user performance.

Pedestrians are able to track the position of a moving vibrotactile target eyes-off, but find it difficult to feel steadily increasing or decreasing changes in intensity. Intensity differences between feedback zones—if even necessary—should be clearly distinct.

Fusion People are able to fuse the guidance offered by directional vibrotactile feedback with their own initiative and instincts. This information can then be used as part of the process of making appropriate navigational choices manually, rather than relying excessively on a device's automated guidance. Future designs should allow use of a device without the need for visual attention, allowing users to focus on their surroundings.

Chapter Five

Shaping

In this chapter we build upon the findings from our investigations into physically grounded discovering and displaying, turning to look at how eyes-off mobile devices can be used to *shape* people's understanding of, and influence the ways that they interact with and move through physical environments. In the previous chapter we investigated the display of a group's central point as a navigational aid for pedestrians, but did not attempt to guide users in their actual route choices. Here we use vibrotactile feedback to allow users the freedom to explore their environment where appropriate, but at the same time guide them to their goal. We investigate this approach using both mobile prototypes and realistic simulations, and in doing so are able to show how the eyes-off techniques demonstrated previously in this thesis can be extended to shape the ways people interact with their physical environment.

5.1 Introduction

In the two preceding chapters of this thesis we have investigated eyes-off methods for supporting people in discovering and displaying geolocated digital information. We have concentrated primarily on using mobile devices to find or browse pre-existing digital content that is connected with (or related to) physical objects that are near to the user. But this focus on finding digital content, while effective in our aim to avoid constant eyes-down interaction looking at a screen, could still potentially miss some of the rich non-digital experiences that are possible in and about the world around us. Much of the interest and enjoyment that comes from being in an exciting place often stems from people's immersion in the events, sights and the other people that bring a place to life. Looking purely at discovering or displaying digital representations of these places could leave out some of the important richness of the physical world.

Pedestrian navigation is a prime example of how the use of technological aids can lead to users becoming disconnected from their physical environment. Current systems provide accurate turn-by-turn directions but, as an unintended side-effect, can lead to complete reliance on the device, rather than on senses or environmental cues, creating a class of users who are oblivious of the sights around them, with a single-minded focus on getting to their destination via the route the device dictates. In this chapter we take a different approach, exploring eyes-off pedestrian navigation designs that allow people to personally take command of the route decisions they make, with the system offering tactile feedback to help shape and inform without controlling their interaction with the physical space.

Our first prototype provides directional assistance toward the user's destination in the form of a fixed width, low-resolution vibrotactile feedback zone. In a similar interaction style to our earlier vibrotactile designs, pedestrians can casually scan to discover the direction of their goal using a handheld device, feeling feedback when they point in its direction. No route or distance feedback is given, however; people must use their own initiative to choose routes. The design is intended to provide underlying reassurance that the user can get to their desired destination, but also to allow them flexibility in the choice of route.

Building upon this initial design, our second prototype modifies the approach to consider how the maps used in existing navigation devices might be appropriated to help users realise the potential for exploring the place they are in. The design uses map data to estimate the number of routes in the vicinity of the user's current location and expands or contracts the apparent size of the feedback area accordingly. This and our first prototype are evaluated first by modelling in a realistic simulator (adapted from that in the previous chapter), and then compared against each other in a large area user study. The results validate the simulation's predictions, helping to illustrate the benefits of both prototypes, and showing the potential of our second prototype for supporting serendipitous exploration.

In the second part of this chapter we turn to consider how it might be possible to construct different representations of physical spaces, by providing navigational feedback that represents the experiences of others in the surrounding area. Our earlier prototype in Chapter 4 displayed the central point of a group of people navigating within the same physical area. Here, we investigate techniques for using location data from people who are *outside* of the navigation process. Our third prototype offers three separate feedback methods, adjusting the feedback given to reflect the most shared or most popular locations where content has been published on social networks.

This third prototype is simulated using the same modelled environment as our earlier designs, and in the process shows how geolocated social media could be used as a novel navigation aid. The evaluations of this and our first two prototypes allow us to measure the success of our designs for nudging and shaping the navigation behaviour of pedestrians. From these results we are able to further extend our design recommendations for eyes-off physically grounded interaction.

5.1.1 Shaping physical interactions

The following scenario illustrates how the routes Alex—who used his mobile in previous chapters to discover and display content—chooses to take through a space might be influenced by his device in a more physically-grounded manner, but without restricting or controlling the way he interacts with his environment.

> *Alex is visiting Rome for the first time, and is looking forward to meeting some friends at a good local restaurant. Taking out his mobile device, he sees the arranged meeting place is about 2 km away. It's such a lovely spring day, so, with time to spare, he roams freely in the rough direction of his goal, taking in the maze of alleys and quirky shops all around him. After about 10 minutes he scans left to right with the phone; the device vibrates to reassure him he's still on course, and its wide feedback arc indicates that there are many possible routes to his destination. It feels good finding his own way, so he continues to make his own choices, enjoying the area around him. A little later, he comes to a main junction. Should he turn left or right? He'd better get this right, he thinks. Scanning again with his mobile, the vibration feedback is now more targeted, and he walks on with confidence . . .*

This scenario illustrates the physically-grounded shaping of people's behaviour that we examine in this chapter. Alex is free to navigate throughout the city, but his mobile device offers low-resolution feedback on demand to help him walk in approximately the right general direction. Furthermore, the feedback the device gives him varies based upon the properties of the environment he is in at the time. In the scenario, Alex is aware that there are alternative routes available to explore, with feedback based upon the number of paths around him. Here we investigate several approaches for estimating the possibility for exploration, first by using actual location maps, and then by using social media updates and publicly-shared geotagged digital content as surrogates for route information.

Pedestrian navigation has traditionally consisted primarily of simple physical cues: maps, signs or compasses; or physical interactions, such as asking strangers for directions. Since the emergence of GPS-capable mobiles, however, handheld digital navigation tools have improved greatly, allowing pedestrian navigation to be simplified even further. At the tap of a button these increasingly-ubiquitous devices are able to calculate the ideal route from A to B almost instantly, guiding people with directions between waypoints and helping out if they stray from the quickest possible path. With these directions always to-hand on the devices we constantly carry, it is easy to be reassured that we now never need to worry about being lost, taking the wrong path, or losing our bearings in an unfamiliar place.

In reality, users of pedestrian navigation systems are often *completely* lost, in that if their device stopped working, they would not know where they were, or what to do next [19, 20]. In this sense, the navigation support available on the device is often more of a crutch that people become dependent on, rather than a liberating feature. It has also been shown that such systems affect how people learn to navigate independently [4, 83], and make them less able to guide others in the area.

Instead of looking around and enjoying their surroundings, people using current pedestrian navigation systems are prompted to speed directly to their goal, heads-down checking a display or listening to headphones for the latest instructions about where to turn next. Our goal in this chapter is to remove some of this division of attention between the device and the real world it describes. We aim to prompt people to fuse their view and knowledge of a location with the feedback given, empowering users to find their own way, trying to free them from the need to constantly look at a screen or be passive, micro-managed agents, following turn-by-turn instructions. Instead we allow them the flexibility to actively take control, and wander where their imagination takes them, with no need to worry about getting back to the 'correct' path.

Our approach recognises that pedestrian navigation, unlike navigation on roads, is often an exploratory process, taking place in semi-familiar areas, or in localities such as towns and cities which are often laid out in a common pattern. Consequently, in the prototypes in this chapter we remove the complexity and restrictions of direction following, aiming to allow people to make their own route choices – instead we offer tactile feedback for the direction of the destination only.

A key component of our approach is a re-envisioning of the 'maps' we use while navigating – here we begin to show how pedestrian navigation might be possible with alternative representations of the locations that the user is navigating through. We envisage this navigation approach not as a replacement for, but as a complement to current systems. Of course, in many situations, it is important to travel in the quickest time or via the shortest path. But we argue that in situations where these aspects are not particularly important, users might find it more enjoyable to turn off the demands of instruction following. Instead, we imagine people wandering through and impulsively exploring the interesting places around them, with occasional reassurance that they're heading in the right direction.

5.2 Exploring dynamic navigation

For many current navigation applications, a crucial requirement is a detailed map of the area, whether in digital, physical or mental form, necessary in order to be able to plan routes between waypoints. However, while these planned routes allow people to get directly to their goal, an eyes-off approach that prompts people to make their own route choices might offer benefits in allowing spontaneous exploration or even just in choosing the most personally preferable of multiple path options. Indeed, while location maps are widely available for road networks and, increasingly, pedestrianised or indoor environments, we argue that a system offering a broader navigational cue can be of use in the many places that may never be fully digitalised – consider navigation through a live music festival, wandering in open parkland, and many places in developing regions.

We created two prototypes to investigate eyes-off physically-grounded pedestrian navigation via vibrotactile feedback. Previous work has investigated the use of a general directional haptic cue for situations such as bike-based tourism [114], but

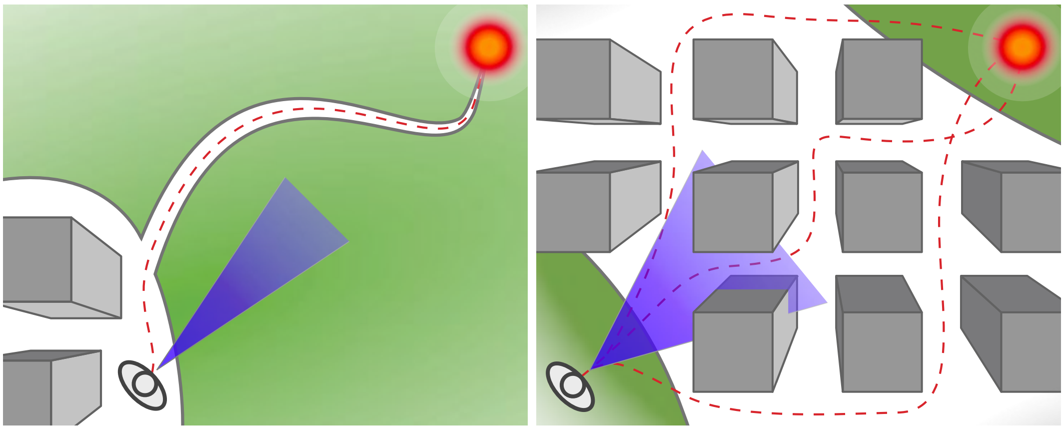


FIGURE 5.1: Left: *static feedback*. The system provides a constant-size feedback arc as directional assistance (shown in blue) to help users find their goal, but no indication of the number of routes available, or their direction. Left and right: *dynamic feedback* is directly related to the available path options. When fewer routes are available (left) the feedback area is small, expanding when there is more choice (right). In both systems the centre of the feedback area aims directly at the goal.

in our approach we apply the technique to navigation while walking and, in our second prototype, offer a richer vibrotactile cue to promote exploration. We build upon the results of our social navigation experiment in Chapter 4, extending this concept in our second prototype to allow users to get a sense of the path choices around them. This design aims to offer the user more freedom where appropriate, giving an indication of the degree of choice available, and providing opportunities for off-the-beaten-track exploration without restricting route decisions. Unlike some previous approaches ([159], for example, which provided feedback varying as function of possible paths) we do not give any indication of the distance to the target, focusing instead on the benefits of giving users familiar and always-available reassurance that they are heading in the right direction.

5.2.1 Prototype designs

Our first prototype—*static feedback*—uses a fixed angular width for the feedback given, relying on the user to observe potential route options themselves, and make appropriate path choices (see left half of Fig. 5.1). The second prototype—*dynamic feedback*—varies the width of the vibrotactile area to give more information about the user's immediate environment (see Fig. 5.1, left and right). By incorporating this extra aspect, pedestrians are able to sense whether alternative routes are available, but are still free to pick their own path at any point.

Our prototype systems are implemented using Nokia N95 mobile phones. For feedback and device movement sensing we use a SHAKE[1] sensor pack. Positioning information is provided by the N95's onboard GPS receiver. For this concept prototype the phone is attached to a lanyard worn around the user's neck, and the SHAKE is held separately in-hand, enclosed in a dummy mobile phone. This was a design compromise chosen to minimise cross-device sensor interference while still providing a realistically-sized object that users could comfortably hold to feel for feedback.

Previous research has shown how feedback can be a function of possible paths through the environment [159]. Our prototype designs were inspired by this finding, and the results from evaluation of our social navigation prototype in the previous chapter, which highlighted that vibrotactile angular widths need not be particularly small or precise as in our earlier prototypes – indeed larger angular widths help to minimise user frustration, and have surprisingly minimal effects on user performance. Our systems use the same minimum angular width of 60° as previously, with the static feedback system using this at all times.

Dynamic feedback

The dynamic feedback approach alters the target width based on the number of potential alternative paths found. While using the prototype, alternative route candidates are calculated by testing for paths from points in the area directly in front of the user's recent trajectory. The number of alternative routes is then used to directly resize the feedback area. Routes that add more than 25 % to the distance of the shortest path from the user's current location are discarded. The number of paths remaining is used to directly resize the feedback area, but this is limited to a maximum of 120° to avoid the excessively long routes that might result from users following the edges of the feedback zone.

For this prototype we precomputed a shortest path matrix using the Floyd-Warshall algorithm [46] from a graph of routes in the local area. This step allows the prototype to perform realtime pathfinding calculations; for a real device we imagine that this data would be server-based and retrieved on demand.

---

[1] See Chapter 2, Section 2.5 for more details.

Simulating the systems

Previous work by Williamson et al. [161] (discussed briefly in the previous chapter) has demonstrated the benefits of using simulations for the design and initial testing of interactive mobile prototypes. Specifically, relatively simple simulations can allow quicker and cheaper testing and refinement of complicated mobile systems, and are capable of accurately modelling both basic user behaviours and complex external uncertainties.

We adapted the custom-built simulator described in Chapter 4 to model both prototypes, allowing refinement of their designs without the need for multiple field trials. The design of the dynamic system was determined through an iterative simulation process, rather than multiple preliminary field trials, allowing us to modify its behaviour and then test the effects on the movements of the simulated users. The parameters for the simulator were estimated based on the same simple assumptions about human movement and the way we expected people to use the system.

We created an accurate model of the planned field trial environment, and used this for 500 simulated runs of both the static and dynamic feedback prototypes between the same start and end points as those used in our trial (detailed in the next section). The simulator predicted mean completion times of: 20 min 52 s (s.d. 7 min 13 s) for the static system, and 20 min 33 s (s.d. 7 min 19 s) for the dynamic system. Mean distances walked were 1.32 km (s.d. 0.42 km) and 1.30 km (s.d. 0.42 km) for the static and dynamic systems, respectively. Simulations were run with Gaussian GPS noise (s.d. 8 m) and Gaussian angular noise (s.d. 8°).

We chose to increase the simulated walking speed in response to the results of our experiment in the previous chapter, which showed that participants had walked slightly faster than expected. In this simulation the walking speed was 1.0 m/s (±0.2 m/s), or 3.6 km/h (±0.72 km/h), checking for feedback every 30 s.

5.2.2 Experiment: shaping pedestrian navigation

We conducted a field study to investigate the static and dynamic systems' effectiveness in a realistic navigation scenario, and to validate our design simulations against the real-world results. The design of our trial was based on methods and recommendations from previous assessments of performance with mobile navigations devices, including both field studies (such as [50] and [164]), and laboratory experiments (such as [124]).

We had two overall research questions. Firstly, we aimed to test the viability of the haptic navigation approach beyond that explored in the previous chapter: can pedestrians navigate to a destination knowing only its general direction? Secondly, focusing on the exploration and freedom offered by the system: does the dynamic feedback prototype's coupling of feedback size to path variance have an impact on users' exploration of their surroundings while navigating?

Method

After an initial pilot study with four participants, 24 participants aged from 18 to 65 were recruited for individual trials to help understand potential usage of the system. Fourteen participants were female, ten were male; 13 were members of university staff, 11 were students. None of the participants worked in areas directly related to HCI.

Before the study, each participant was randomly assigned to use one of the two prototypes, in a between-subjects design. Fixed start and end points were chosen at the edges of the approximately 0.5 km^2 study area. The study area comprised of both university campus and open parkland, in order to give participants exposure to navigation with the system through both urban and rural areas. The straight-line distance between start and end points was 0.77 km, and the shortest walking route (when keeping to paths) was approximately 1 km. These well-spaced points allowed us to measure participant performance at a much greater distance than that commonly used between pedestrian turn-by-turn waypoints, especially in urban environments.

At the start of each study session participants were met individually and given an introduction to the concept and basic usage of the system they would be using, followed by a short demonstration of the prototype. After a brief training session (less than 30 s per user) in which they felt example feedback, participants were led to the pre-determined starting point on campus (see point A in Fig. 5.2). When at the starting point, they began using the system to scan for and attempt to navigate to the end point (point B, Fig. 5.2).

No description or guidance about the location of the end point was given to participants, (neither before nor during the study), in an attempt to minimise potential effects from participants' prior knowledge of routes to the goal. While navigating, participants were free to take any route they wished over the entire study area, while the researcher followed and observed. Upon reaching the end point a short semi-structured interview was conducted to gather opinions and experiences, and all participants were rewarded with a bookstore gift voucher as a token of appreciation.

Measures

In addition to participants' comments and opinions in the post-study interviews, and notes of observations from each trial, we collected detailed device logs allowing in-depth analysis of participant behaviours against our research questions.

Viability We measured the success of the system as the overall percentage of participants who found their way to the end point. Viability is reinforced by participant observations and remarks, and by looking closely at walking speeds, specifically the variance over the trial and the amount of stopping required.

Freedom The freedom offered by the dynamic feedback prototype is measured by comparing the variation in paths taken by participants over both systems. In addition, comparison to results from our static feedback prototype allows a measure of the extra distance and time cost of any exploratory behaviour.

5.2.3 Findings

All participants successfully completed the navigation task and found the end point with only the vibrotactile feedback to guide them. Table 5.1 shows the results for the time taken and distance walked for both systems.

Measurement (units)	Static		Dynamic	
Time taken (min:s)	19:02	(5:36)	17:24	(5:25)
Distance walked (km)	1.65	(0.58)	1.53	(0.39)

Table 5.1: Means and standard deviations (in parentheses) of the distances walked and times taken to navigate from the start to the end point using the static and dynamic navigation systems.

The distances walked ranged from 0.97 km to 2.39 km for dynamic feedback and 1.08 km to 2.93 km for static feedback. Times and distances were not significantly different between feedback types (ANOVA, time: $p = 0.5$; distance: $p = 0.59$). Clearly, users were able to navigate to the end point without the need for turn-by-turn guidance, and over a distance twice as far, on average, as that demonstrated in the previous chapter. The mean times taken and distances walked are longer than those for the shortest path, but the ranges of times, distances and routes taken (see Fig. 5.2) suggests that this has been as a result of the variance in path choices.

Path choices and walking speeds

Figure 5.2 shows the routes taken by participants using each prototype, and also the shortest route when keeping to paths – the likely choice of a turn-by-turn navigation system. Interestingly, although the destination point was the same for both systems, many participants using the dynamic feedback tended to stick more closely to the main thoroughfare of the university campus, while those using the static feedback often took a less well-trodden route. This suggests that the varying vibration has allowed users to combine the feedback given by the system with both the path cues in their immediate environment and any prior knowledge of the area. Participants using the static feedback seem to have felt obliged to follow the target direction more closely.

Using methods adapted from [33] for gait phase analysis, we can again look at participant walking behaviour in detail, as shown in Fig. 5.3. Looking first at walking speeds (top left), there was very little difference between systems, though those using the static feedback had a slight tendency to walk faster. This similarity in walking speeds was shown throughout the task (top right), regardless of the distance to the goal. When looking at walking rates in conjunction with the feedback given (bottom), we can see that participants using the dynamic feedback probed for feedback for a larger proportion of the time.

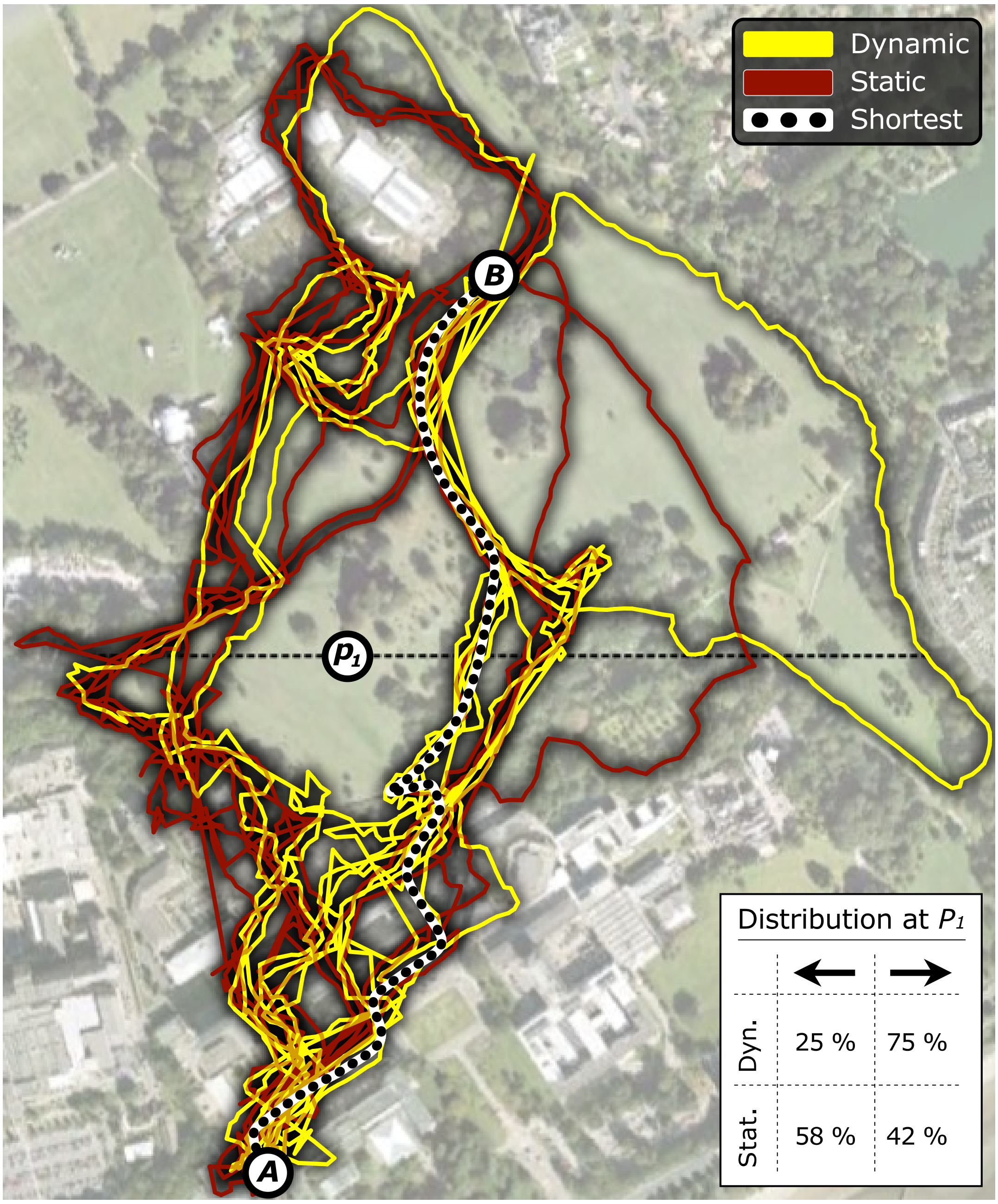


Figure 5.2: Routes taken from points *A* to *B* by all 24 participants during the study; shortest path overlaid. Inset: distribution between main routes. Participants feeling dynamic feedback tended toward the main campus thoroughfare; those feeling static feedback often took less familiar routes.

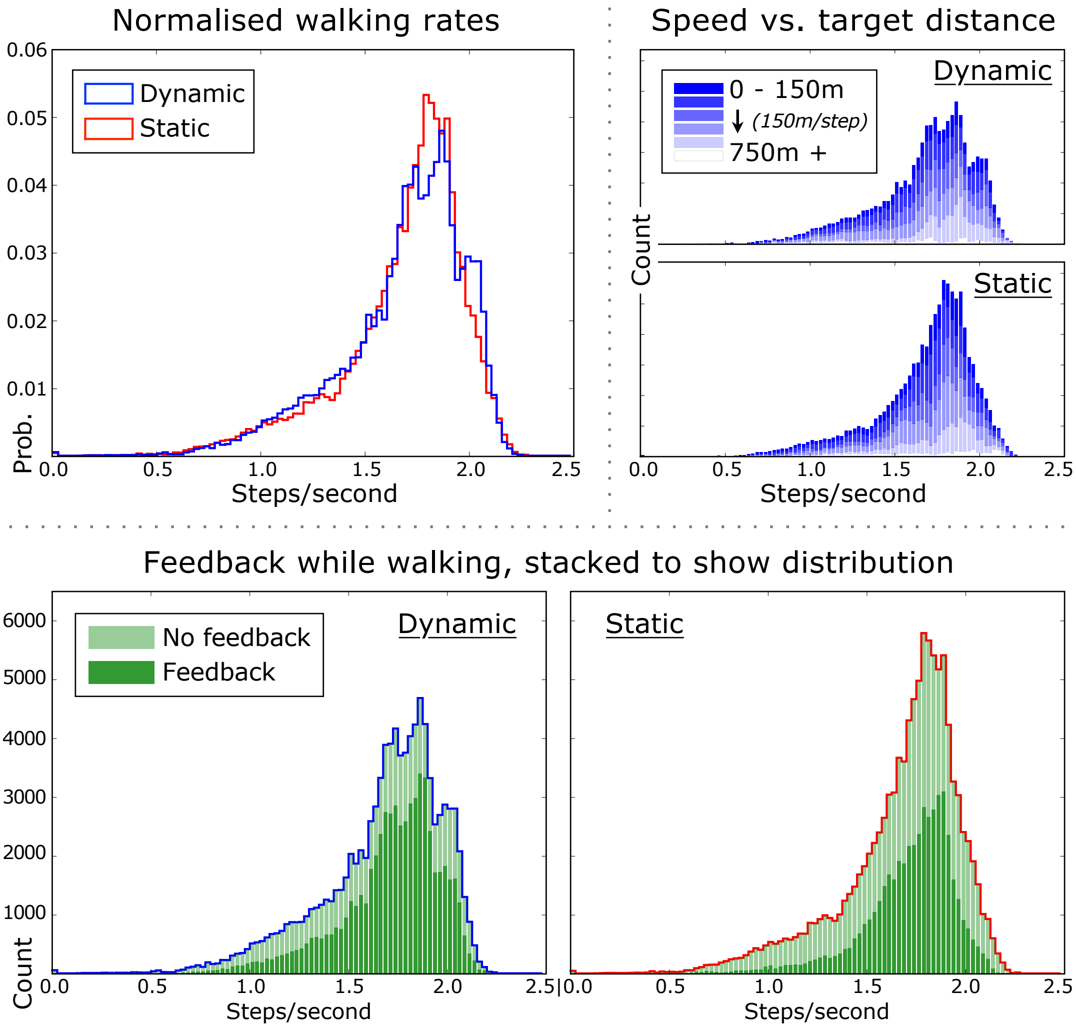


FIGURE 5.3: Top left: Walking speeds for each system: overall speeds are clearly similar. Top right: Walking speeds for 150 m segments of the routes taken. Similar walking rates were maintained throughout the task for both systems. Bottom: Walking speeds while feedback was activated. All participants walked and interacted simultaneously, but those using the dynamic feedback system interacted more, proportionally

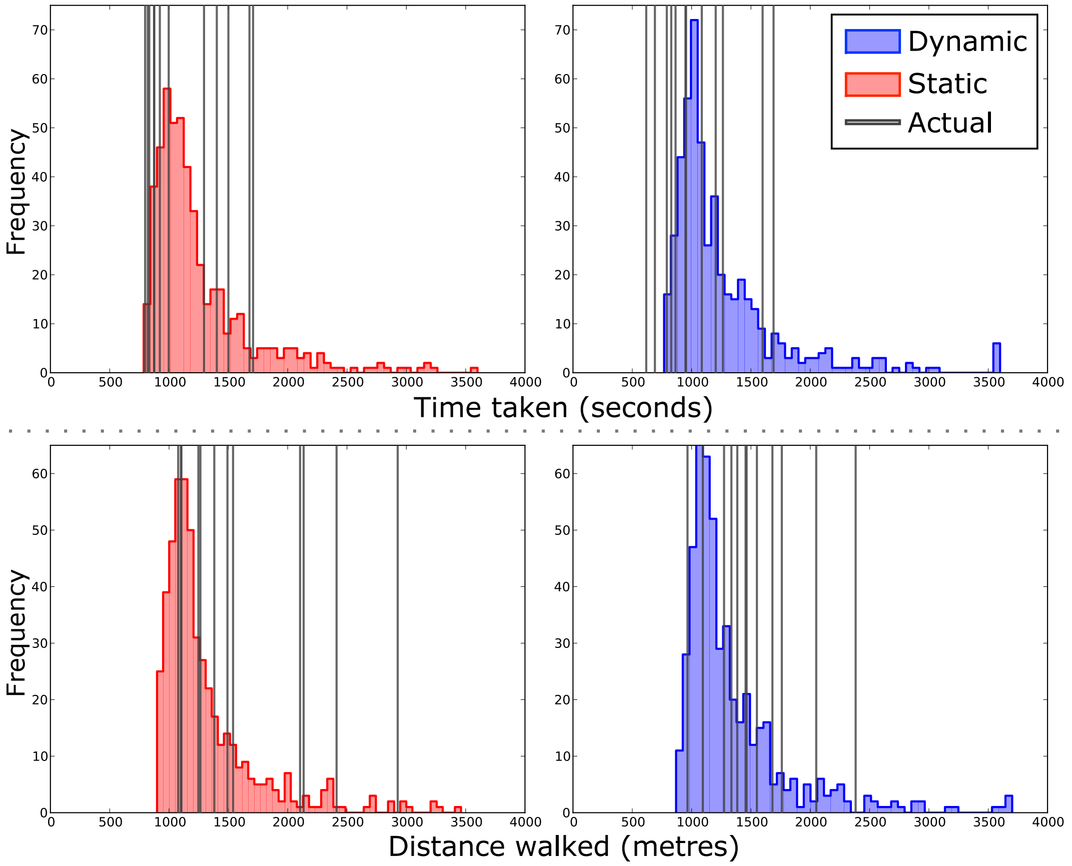


Figure 5.4: Simulated times taken and distances walked for each prototype (shaded areas; distribution after 500 runs). Actual results from study participants are shown as vertical lines. The limited pathfinding ability of the simulated users has caused long tails, but most study results fall well within the range predicted before the trial.

Comparison to simulated routes

Figure 5.4 compares results from the simulator with those from the field study. Times taken and distances walked in the trial are within the ranges predicted in the simulation, although in the study participants have walked slightly faster than the 1.0 m/s (3.6 km/h) we assumed during the design of the system.

Interestingly, time and distance differences between the two prototypes in the field trial results were not present in the simulated runs. In the field trial, participants using the dynamic feedback reached the goal faster and walked shorter distances, on average. These differences are likely to be a result of the limited walking behaviour model in the simulator – there are many external factors that we were unable to model, such as the possibility that participants have some level of prior knowledge of the study environment layout, which may have subsequently influenced their walking behaviour. Despite this, however, the times and distances measured in our field study lie well within the range predicted by the simulator, highlighting the value of simulations such as these for predicting the effectiveness of interactive mobile navigation systems.

Participant observations and feedback

Observations validated the analysis of device logs, confirming that participants walked at a steady pace for most of their route, with some making occasional pauses to double-check for confirmation at major path junctions. All participants held the prototype by their side for the majority of the task, tending to look ahead or around their location, rather than at the device. All participants except for one enjoyed using the systems, and many were surprised at their effectiveness despite some initial scepticism. Several participants remarked on the ability, as articulated by one participant, to *"combine technology and knowledge of the environment to pick the right path,"* and that as they were in no hurry it was *"good to be able to explore."* Three participants said that despite their success they would not use haptics for navigation, however, because they preferred to have constant knowledge of their position and destination.

The participant who disliked using the system did not like holding the device constantly, but would have liked to repeat the trial with the device kept in a pocket to be used for occasional route updates. Half of the participants using the dynamic system explicitly commented that they liked the varying feedback, finding it helpful to know when they could take a different route. This behaviour seems to be reflected in their route choices, with one of these in particular taking an extended route to *"go down here because I've never been here before."* Most participants suggested potential usage scenarios for this low-attention method of navigation, ranging from searching for catering venues to simple, low-cost tourist guides.

5.2.4 Discussion

All participants were able to find an unknown target location with only directional vibrotactile feedback as a guide. The lack of turn-by-turn navigation guidance did not have a noticeable effect on walking behaviour, with brief pauses to confirm bearings being the only times participants stopped for the majority of routes. Users were required to choose their own path to the goal without a map of the area or turn-by-turn guidance, but in all cases they were able to navigate without needing to stop or backtrack unnecessarily. As can be seen in the walking data (and confirmed by participant observations) users preferred to keep track of the target direction most of the time, but were able to do this casually and without sacrificing attention. This eyes-off approach to using the system was unprompted, and chosen by participants naturally, demonstrating the success of the systems at allowing physically-grounded interaction without unnecessarily interfering in users' normal behaviour.

While it is probable that the lack of waypoints might be an issue for navigation over much larger areas, users have had no trouble selecting appropriate routes over distances averaging at least 1.5 km. The variation in paths between users of the two systems shows interesting behaviours around commonly-travelled areas. One example of this is the divergence in paths around point P_1 in Fig. 5.2, where participants using the static prototype more often chose one route option, and those using the dynamic system chose another. The fixed feedback of the static design has perhaps led participants to believe that there was only a single 'correct' route. Conversely, the variable feedback of the dynamic system allowed these participants the freedom to choose from the variety of paths available at an earlier decision point.

We aimed to allow users more autonomy in route-finding while still being able to navigate to a target, and this freedom is evident to some extent in the range of paths taken. Many participants chose to follow familiar paths when given the option, but, interestingly, some outliers took the opportunity to explore an area they were not familiar with. Most participants kept to major paved paths while in a rural environment, but some, using either system, decided to take more direct routes (over wet parkland) when possible. This is an interesting behaviour (and not an aspect emphasised in our design process), though we suspect that users might prefer actual paths when navigating to self-selected destinations. In some cases these shortcuts have caused the participant to reach a dead end – this is an example of where our design does not offer the precision of a traditional turn-by-turn navigation system. However, even in these cases users have managed to find the end point with no further assistance.

There is no statistically significant evidence of added costs in user performance while using variable tactile feedback; in fact, on average, users of the dynamic system found the target more quickly and in a shorter distance. The less-precise dynamic feedback did not adversely affect users' navigation abilities, suggesting that this new approach to pedestrian navigation could effectively complement existing turn-by-turn or static haptic methods where appropriate. Participants offered positive feedback about both systems, with some explicitly commenting on the potential for exploration, particularly in tourist scenarios.

Simulations of user behaviour prior to the trial showed behaviours largely similar to those demonstrated by participants during the experiment, highlighting the benefit of simulations for evaluating mobile navigation device usage.[2] A straightforward reproduction of the study area, combined with a basic model of walking behaviour, incorporating the complexities in positional and sensing hardware, has allowed accurate predictions of the results of a real-world field study. The demonstrated accuracy of the simulations highlights how it is possible to evaluate such devices with real-world accuracy without the need for complex and costly field trials.

Our dynamic prototype used a matrix of potential paths to offer users location-based feedback on demand. But the paths available through a place are just one of the properties that might be incorporated into people's route choices while walking. Other factors could include particularly interesting-looking locations, or attractions that are known to be highly-rated. While existing navigation systems do incorporate this information to some extent to indicate well-known points of interest, there is a wealth of contextual data about regularly visited places and talked-about attractions in the content that people share on social networks. In the next section we investigate the use of social media and publicly-shared content for navigation. These are, of course, similar types of content to those we used in our earlier chapters, but here we take a different approach, using this media to shape people's movements, rather than for direct discovery or display.

5.3 From dynamic to socially-shaped navigation

The term 'social navigation' most commonly refers to navigation shared between multiple people through digital spaces [100]. Here, though, we consider digital social navigation applied to *physical* spaces. While maps and guides offer potential

---

[2] Recent research ([98], for example), has built upon this and our previous work, evaluating several aspects of directional vibration for navigation, and helping to validate the simulations used here.

routes and pathways through a location, geolocated social data can indicate positions where people have physically been, with quantity as a straightforward measure of the popularity of a particular place.

As exposure of people's everyday experiences online increasingly becomes the norm, many of the most popular social networking services are adding functionality to allow users to pair location data with their updates. The location information from these status events ranges in detail from approximate city-level data to precise latitude-longitude coordinates, depending on the available positioning granularity and the level of privacy chosen by the author. Many social media services now provide public APIs that allow filtering and retrieval of these social updates by some precise location.

We investigated the navigation possibilities afforded by the location information that is paired with the huge quantities of geolocated content generated and publicly shared each day, worldwide. While our previous social navigation work in Chapter 4 used positioning data from co-located users moving in cooperation, here we are able to incorporate the location of people external to the navigation process, both location and, quite likely, in time. In a similar way to the approach used by Jones et al. [74], the places in which strangers were spurred to share their media, thoughts and opinions can be used to inform and shape the user's behaviour.

In this section we demonstrate the use of content shared on social media websites for ad-hoc estimation of the route variability in a particular location, using both the extent and distribution of social network updates as a proxy for detailed map data. As demonstrated in the field trials of the two prototypes described earlier in this chapter, using simulations for the design and evaluation of interactive mobile systems can be both accurate and reliable. Accordingly, rather than trialling physical social media navigation prototypes, we created simulations of several methods, allowing comprehensive evaluation of their potential effectiveness without the need for additional user testing at this stage.

5.3.1 Prototype designs

The distribution of social media updates is usually irregular and difficult to predict far in advance, with updates posted spontaneously, often concentrated around places that are popular or newsworthy at a particular instance in time. However, when using

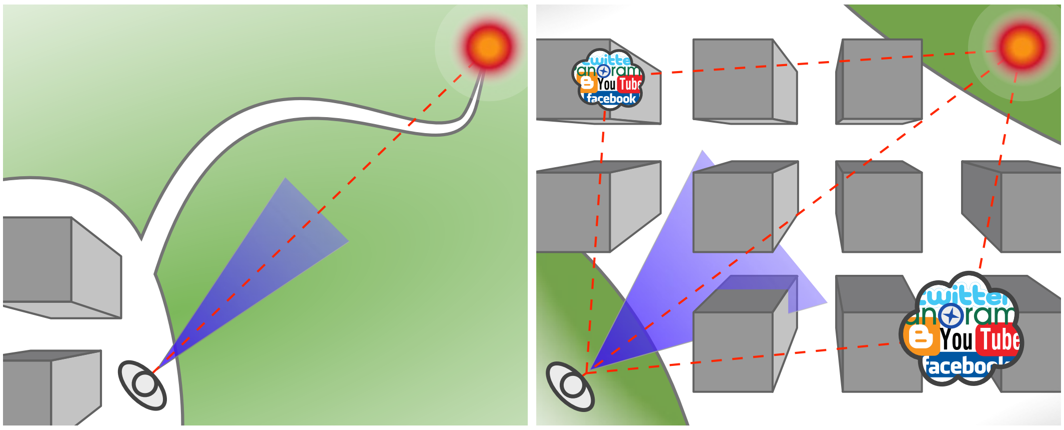


Figure 5.5: Our socially-shaped navigation approach links shared geotagged social media data to potential route options. Left: when little or no content is available, the feedback area (shown in blue) is small, aiming directly at the goal as in our earlier static navigation prototype. Right: when more content is available, the feedback area expands to indicate that this location is often shared, and so may be worth exploring. No indication of actual route availability is given, however.

social media for navigation this dispersion gives the data extra value. By constructing several unique views of publicly-shared geolocated content, we are able to offer the possibility for custom social tours through public spaces. Choosing pictures might help the user explore the most scenic parts of a park or nature reserve, while the use of social network status updates could hint at the most popular places during a live music festival. Although clearly inferior in absolute accuracy to location maps, we believe this approach offers particular benefits in those scenarios where maps struggle to keep up with changes in scenery (such as live events), or when context is important and, for example, a person walking alone might take a different route than when walking with friends.

By retrieving these located posts around a navigating user, we are able to build on our previous social navigation work to construct a social representation of a pedestrian's surroundings. Using the same variable feedback approach as our previous dynamic navigation design, the positions of social media (instead of route possibilities) from Twitter and Panoramio[3] were used to predict potential 'paths' between the user's current position and the goal (see Fig. 5.5). Media locations that added more than 25 % to the direct distance between the navigator and the goal were again discarded. The number of routes from the user's location through each permutation of social media items was used to expand and contract the feedback area, using the same minimum (60°) and maximum (120°) angular widths as previously. In addition,

---

[3] http://www.twitter.com • http://www.panoramio.com

we also took into consideration the distance of the user from the location of the social content in order to prevent, for example, content items clustered around the goal from affecting the apparent path choices throughout the entire journey.

We designed three separate approaches, creating interfaces for each in our simulator. The designs were chosen in order to investigate how different sources and interpretations of dynamic social content might influence the behaviour of pedestrian navigators, helping them explore particular views of their surroundings.

Nearby social media

Our first design uses the public Twitter API to retrieve status updates posted in the area between the user and their goal. Due to a lack of sufficient realtime social network data in the modelled study area, however, we collected an aggregated record of Twitter posts over a two-week period to give a more accurate picture of the spread of social updates in the area.

Social media hotspots

For the second design we use the same aggregated Twitter data to estimate route possibilities ahead of the user. However, in this design we group results into hotspots based on the quantity of updates in a single location. Areas with at least three updates within 100 m are considered a cluster; those with fewer are ignored and do not influence the feedback area. Using clustered results, rather than every update available, allows us to investigate how the features of the navigation area—in this case, place popularity—rather than its paths, might be used to aid navigation.

Geolocated images

Our third design uses the Panoramio API to retrieve images in the user's vicinity, rather than social updates, using these in the same way to adjust the width of the feedback area. As geolocated images shared on Panoramio are less time-dependent than Twitter updates, there is no need to collect images over a longer period of time; instead, we retrieve images when needed, simulating the possible operation mode of a real-world system.

5.3.2 Experiment: simulating socially-influenced navigation

Five-hundred iterations of navigations using each of our designs were simulated using the same parameters as used for our initial prototypes, with the same modelled area, and identical start and end points. Although the results from our first study suggested that the simulated walking speed chosen was slightly lower than pedestrians' actual speed (when compared to real-world walking data from the trial), we chose to keep the same 1.0 m/s (±0.2 m/s), or 3.6 km/h (±0.72 km/h), to allow comparison between results from both simulations.

5.3.3 Findings

Table 5.2 shows the resulting times taken and distances walked for each of our designs. The simulated behaviours using social media data are in line with the results from our earlier simulations, indicating that this simple use of geolocated social media updates might be a viable source of 'map' data, especially when location models are not available.

Differences in navigation behaviours are evident when we look in detail at the routes taken by simulated users. Figure 5.6 shows the paths during the simulations for each social navigation method. When comparing routes in the built-up area there are few variations between the three designs. However, routes taken in open parkland are clearly more widely distributed. While both clustered and unclustered social media updates have prompted simulated users to take largely similar routes, the large quantity of images in the open area (resulting in a larger feedback width) seems to have caused more variety in routes taken.

System	Distance walked, km		Time taken, min:s	
Nearby social media	1.36	(0.54)	21:22	(9:11)
Social media hotspots	1.29	(0.40)	20:09	(7:01)
Geolocated images	1.34	(0.45)	20:48	(7:38)

TABLE 5.2: Means and standard deviations (in parentheses) of the simulated distances walked and times taken to navigate from the start to the end point for each of the social navigation designs.

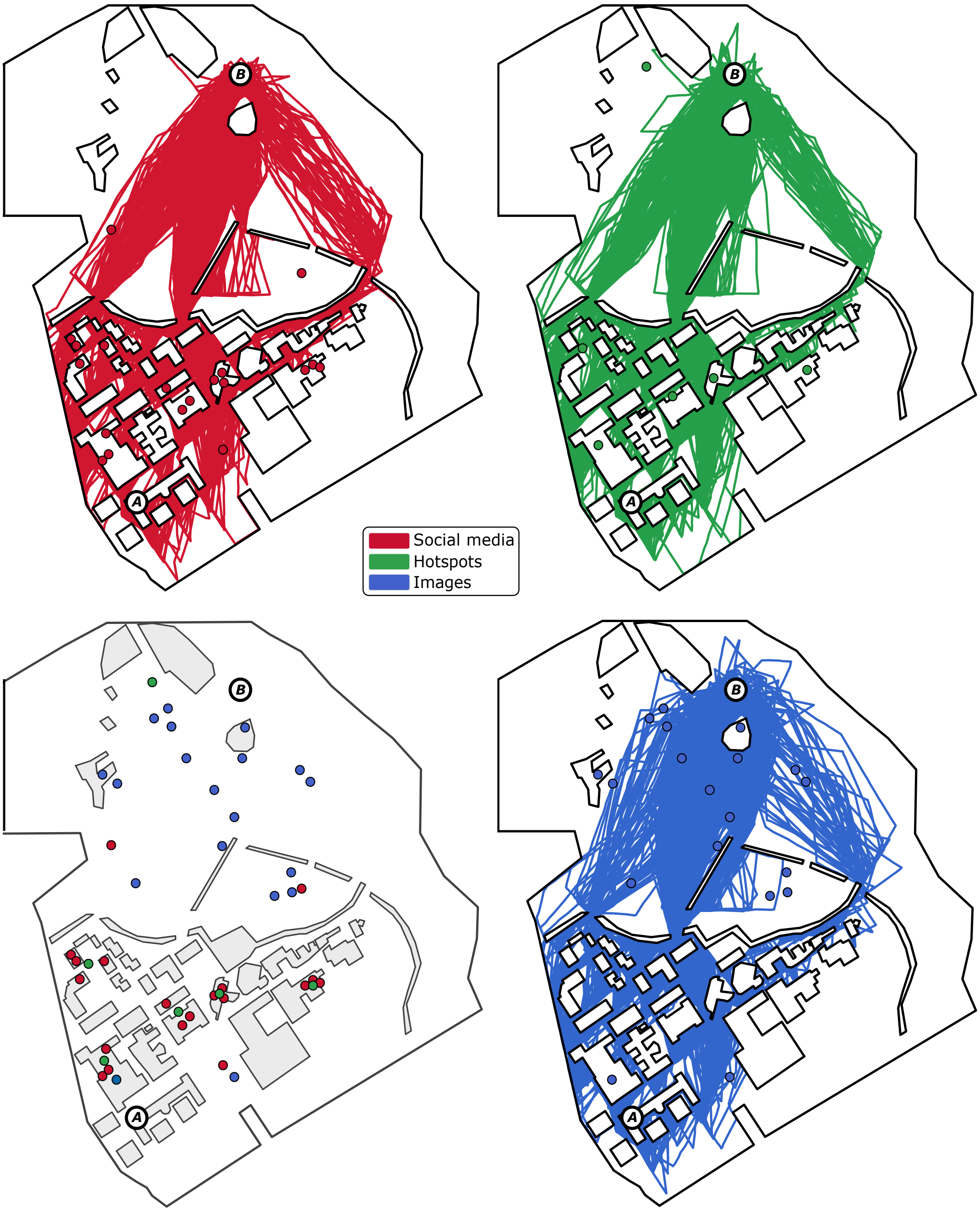


Figure 5.6: Simulated social navigation, showing the routes taken by 500 simulated users for each social navigation method between points A and B. Lower left: the model of the study area used for the simulations with social media positions highlighted for comparison between methods.

5.3.4 Discussion

The use of geolocated images and public social media updates shows potential for allowing social navigation of pedestrianised areas, as demonstrated in the simulations here. The results from this socially-influenced dynamic navigation approach are comparable with those for our earlier simulation of route-based dynamic navigation, which themselves are comparable to the results from our initial field study.

Looking more closely at the simulation results, interesting behaviours are evident when considering the source of the geolocated data. In the trial area chosen for our simulation, Twitter postings are most common around built-up areas, while Panoramio images are more widespread in the surrounding parkland. The spread of content, particularly in the open area, seems to have led to more diverse routes. Although this is perhaps a predictable result, it also hints at the potential for automatic off-the-beaten-track tours of both urban and rural areas, augmented with appropriate social media data, but with no need for manual route generation. Indeed, while our socially-influenced navigation designs are, of course, simplistic methods for estimation of route variation and location popularity, it is relatively easy to create custom views of this content, closely tailored to particular user or organisational needs.

Although in these simulations we used aggregated data, realtime social media retrieval is possible in many of the most widely-shared locations, such as busy cities or live events. Where our approach offers most benefits is, we argue, in its direct transformation of the social popularity of an area into navigational support given to a pedestrian navigating through it, regardless of the quantity of updates that are being posted. Simple modifications to our methods could use the area maximum to calibrate the local quantity of social updates, providing an instant picture of currently evolving events, and helping to alert users to the possibilities around them.

5.4 Conclusions

In this chapter we have explored and evaluated eyes-off directional vibrotactile feedback for shaping pedestrian navigation. In our first prototype, haptic feedback was used to support navigation without the need for waypoints or turn-by-turn directions. Our second design extended the technique to offer explicit support for exploratory navigational behaviour. In the final part of this chapter we simulated several designs for socially-influenced pedestrian navigation. The promising results highlight the potential for this type of system in the wild.

Haptic navigation has allowed pedestrians to find their way to their goal, but stimulated them to make sensible route choices, fusing their view of the world around them with the simple directional feedback given by our prototypes. In the field trial of our static and dynamic prototypes, pedestrians using a minimal directional vibrotactile cue successfully navigated to an unknown target destination while dealing with the complexities inherent in pedestrian navigation. Users were able to maintain a steady walking pace throughout the trial, with negligible affects on their normal behaviour. Results from our static prototype using fixed-size vibrotactile feedback support and validate those of our group navigation system in Chapter 4, and those from the more advanced dynamic prototype show the benefits of providing users with alternative path awareness via simple changes to angular feedback.

We have simulated three methods for the use of geolocated public and social network data as a proxy for route and obstacle models. The results from these simulations demonstrate how the use of publicly-shared geotagged content could allow dynamic navigation, personalised in realtime to particular needs via a constantly-evolving social 'map.' Future extensions of this work could look more closely at how these social maps might be created, evaluating on a larger scale using live social data.

5.4.1 Designing to shape eyes-off physically-grounded interaction

The field study described in this chapter shows clearly how the use of directional vibrotactile feedback can support pedestrian navigation. In our design waypoints were not used; for much larger areas this may not be practical. A potential solution to this issue is simply to use waypoints that are within the range demonstrated in our trial, but further simulation and validation are needed to explore whether there are alternatives. Regardless of distance, care is needed to ensure that users are not led along dead-end routes that might, coincidently, align with the direction of the feedback zone. A fusion of our design with traditional turn-by-turn guidance, where appropriate, could address this issue.

Variable-width feedback areas, whether route- or social media-based, can help to prompt participants to make their own choices from the available alternative paths, rather than relying completely, and naïvely, on the device's guidance. While currently only simulated, the use of social media updates shows interesting behaviours and findings for the future use of this information for navigation. Careful consideration needs to be given to the effects of guiding people via social content, however – while we aimed to encourage serendipitous exploration, regular use of the same social data by large numbers of users could lead to exactly the opposite scenario. There are also interesting social issues to be aware of – for example, by posting a large number of updates in a particular area, a malicious user could significantly affect all users of the system navigating via a particular service.

Navigation Low-resolution directional vibrotactile feedback can help pedestrians navigate to their goal without the need to look at the device that is guiding them. Designers of future systems in this area should aim to maintain this eyes-off interaction method where possible, but fall back to more traditional navigation designs where appropriate.

Users in our study were able to walk and use haptic navigational feedback simultaneously, with only brief pauses for confirmation at major junctions. Future designs could consider offering more focused feedback in these zones or, as in our dynamic design, making clear where there are multiple route options.

Instructions The need for precise, instructive and controlling turn-by-turn waypoints is far less evident for pedestrian navigation than it is for traditional, car-based GPS navigation. Our study demonstrated participants' ability to use a waypoint-free design over distances averaging at least 1.5 km.

While this result does not, of course, necessarily imply success over far larger distances, the use of this technique for typical city or urban navigation scenarios, in which the distances are relatively small, seems likely to be beneficial.

Exploration Haptic feedback that varies in width (rather than in location, as in the previous chapter), can help to prompt pedestrians to make their own route choices based on personal preference. As we saw in our earlier navigation design, the participants we observed seem to have been fusing external factors with the feedback offered, rather than relying solely on the device for directional instructions.

Chapter Six

The Next Billion People

Throughout this thesis we have explored the design and evaluation of novel interaction prototypes that have been constructed using high-end, costly and often custom-made hardware. In this chapter we turn to consider how these types of advanced interaction methods could be supported for the billions of people who do not—and most likely never will—have access to this high-end hardware. Consequently, our focus is on augmenting users' existing devices and services to allow the use of eyes-off interaction techniques, rather than distributing custom or specialised tools. Here we investigate one potential approach by augmenting an existing popular voice-based telephone service in rural India. In doing so, we are able to highlight the potential benefits for users, and propose design recommendations for future designs that take this particular approach.

6.1 Introduction

For hundreds of millions of people in developing regions, the mobile phone is the primary—often the only—interactive technology that is available. Mobiles are already pervasively used for calling, but these devices look increasingly set to transform into access terminals for remote information services. However, many of these mobiles are likely to remain relatively dumb-phones with only a low proportion being routinely served by a data connection [149]. Furthermore, the users themselves add additional challenges to the goal of universal access: many have a low level of textual literacy, and their prior exposure to computing technology is often very limited.

In the first three chapters of this thesis we have investigated the sorts of state-of-the-art, future-looking devices commonly studied by HCI researchers. For most people in more developed areas, the hardware we have used in our prototypes will become, or is already, commercially available. But for many people in developing regions this possibility is unlikely – these types of devices, and the data connections required to run them, are simply unaffordable. In this chapter we demonstrate how some of the benefits of the eyes-off interaction methods we have studied in previous chapters can be offered to users of low-end mobile phones that contain none of the sensory or tactile features of the devices we have used previously. The prototype we examine here is implemented as an add-on to a voice-based information service that is accessible via any telephone handset, whether landline, dumb-, feature- or smart-phone.

Existing interaction on the voice service we integrate with is via button presses on a handset or, in some cases, basic speech recognition in a small number of local languages. The use of buttons on a handset requires the user to look at the device to interact – a potentially disruptive and distracting task, particularly when using the service to listen to important information that might need to be used or relayed immediately. Our prototype adds remote, server-based recognition of 'audio gestures,' using the back of the user's handset for eyes-off input. These gestures—tap sequences in the current version—are used to control navigation through the voice service, allowing eyes-off interaction with this popular service using any phone.

Viability trials of our prototype, both in technical evaluations, and over a longer usage period with existing members of the voice service, show the benefits of the approach. Furthermore, there is a demonstrated desire by ordinary users of the service to adopt the system. Our evaluations show how the sorts of high-end interactions we have previously studied might be made available to users of low-end devices, and help to extend our recommendations for future systems that explore this design space.

6.1.1 Eyes-off interaction for impoverished platforms

The following scenario illustrates how Hina, who regularly accesses audio-based information services via her mobile phone, might use remote eyes-off interactions to navigate through and interact with the service. Over the rest of this chapter we examine the design, implementation and subsequent user studies of such a system, as an add-on to the *Spoken Web*.

Hina is walking through the fields she farms, inspecting the crop as she goes. In one area the plants aren't growing very well, so she pauses to take out her mobile, and calls the Spoken Web. When the call connects, Hina hears a list of categories that she can listen to information about. She decides that the question and answer forum will probably be the best place to search, so she taps on the back of her phone to follow the voice link. Hina walks around the plants, checking and comparing to a list of common problems and their symptoms from farmers all over Gujarat. There's no need to take the phone from her ear to navigate; instead, she taps occasionally on the back of the device to fast-forward through sections that seem to describe unrelated problems. Soon Hina finds an informative voice clip from a state agriculture expert about the possible cause, and a simple solution. Hina scratches the back of her phone to mark this section as a bookmark – next time she calls she'll be able to find this useful information easily . . .

The Spoken Web [84] is a collection of interconnected *voice sites* that aims to offer remote and widely-available information access while meeting the challenges of cost, textual literacy, data connections and general technology exposure that are common in many developing regions. Its interactive audio applications provide content on topics such as farming or health information over the public telecom network. Individual voice sites are accessed using any type of phone by dialling unique telephone numbers (analogous to internet URLs).

The service combines automatic speech recognition (ASR) and touch-tone dialling (DTMF – dual-tone multi-frequency codes) to allow people to create and browse through spoken content. Potentially, ASR promises eyes-off, intuitive, low cognitive load interaction with audio content, with the benefit that no base level of literacy or numeracy is required. However, the pauses or cues that are needed to prompt speech input and detection can easily upset the interaction flow, especially for short inputs such as 'yes' or 'no' [109]. Furthermore, the language models and recognisers for many languages in developing regions, although improving, are not yet complete [130]. DTMF key tones are quick to enter on mobile keypads, designed to be unambiguous when recognising, and can offer many possible input sequences. However, because the tones generated are echoed in the phone speaker (confirming input, but drowning out incoming audio), commands and responses are often fragmented. It is also usually necessary for the caller to take the phone away from their ear to see the keypad and enter their response to any input cues, potentially disrupting the interaction flow.

Although both ASR and DTMF allow a level of control and interaction with audio content, we argue that there is still much work to be done in terms of improving the expressiveness and range of interactions. While button-based DTMF input is clearly required to make the initial phone call to a voice site, in our prototype in this chapter we have concentrated on addressing the disruption to the call flow that can occur when using the current methods for browsing the content itself. Our prototype provides an extended interaction method for voice sites that aims to allow callers to smoothly navigate through and control the content they are listening to without having to unnecessarily interrupt its playback by removing the phone from their ear.

Our approach uses simple back-of-device interactions—*audio gestures*—for eyes-off interaction on the phones users already own. Previous work has noted and addressed the problems that can result when callers attempt to use a phone's keypad without taking the phone away from their ear – see [88], for example, which addressed these issues by using a modified keypad on the back of a phone. The back-of-device inputs supported by our system, however, are captured during the normal call flow by the phone's microphone, with no client software or phone modifications required. Our prototype runs on the Spoken Web server to recognise these inputs remotely. Previous research on back-of-device non-speech natural audio input has involved state-of-the-art custom-built devices, and users with high levels of literacy and technology experience. For example, the prototype developed by Murray-Smith et al. [101] used expensive 3D-printing to create a textured surface on which various tactile areas could be scratched or rubbed to produce different interactive responses. In contrast to the intended users of these high-end devices, the majority of the Spoken Web's target audience use relatively low-end mobile phones, so we have focused on providing these additional interaction features without requiring users to own a specialised device.

Our contribution in this chapter, then, is an exploration of the ways that impoverished platforms and their users can be afforded the sorts of advanced, eyes-off interactions that we have previously studied for people living in the more developed world. The approach here is not a panacea; nor is it robustly universal: rather we consider it as an early investigation into the potential for eyes-off interactions in less developed regions. Furthermore, this is not an argument for designing technologies that are specifically focused on 'low-end' users – see, for example, the low uptake of specialised devices such as the Motorola FONE, a device designed specifically for 'poorer' users. Rather, we intend to highlight the rich potential that exists in designing interactions that are particularly focused toward the existing devices owned by people who may not have the means to afford or interact with the latest technology.

6.2 Exploring tap-based interaction with voice services

In order to investigate the potential for supporting eyes-off interactions on low-end devices we developed a back-of-device input recognition prototype as an add-on to the Spoken Web. The *TapBack* system allows callers a richer experience with interactive voice sites by enabling *audio gestures* to be used at any time during a call. By using the back of users' phones as an input surface while a call is in progress, we remove the interruptions of ASR or DTMF input and allow users to keep the phone by their ear throughout the call. Unlike previous back-of-device methods, we use the phone's inbuilt microphone to pick up the sounds generated on the back of its case. These sounds are loud enough to be transferred to the other party in the call but, unlike DTMF tones, are not so loud on the caller's end that they drown out the audio being played.

We built upon previous research into interaction that appropriates a device's surface as an input channel (e.g., [101]), using a fast, lightweight recognition engine similar to that used by Harrison and Hudson [54]. Mobile possibilities for these types of interactions have recently been demonstrated as a commercial prototype[1] that can detect tapping locally via an application on a feature-phone. In contrast with these types of approaches, however, our design also separates the source of the audio gestures (the handset) and the system that interprets them (a server accessed over a normal telephone call). Additionally, our technique requires no modification—hardware or software—to the device that is used for input.

6.2.1 System design

To simplify introducing Spoken Web users to audio gestures we chose to apply tapping recognition to the control of audio playback speed. Previous analyses of voice site

---

[1] *TouchDevice* – see: http://www.inputdynamics.com

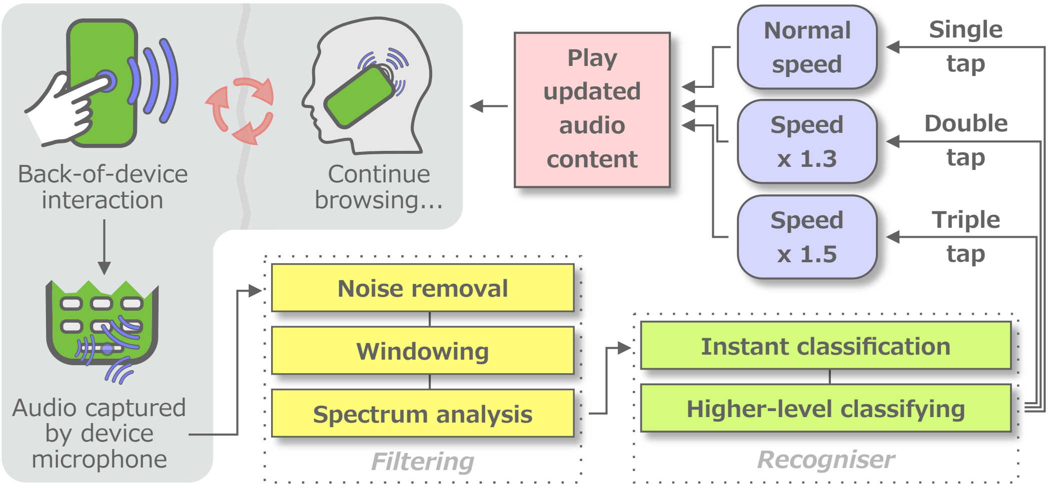


Figure 6.1: Our prototype in operation. Audio gestures are generated by the caller locally on the back of their handset (shaded region), but analysed remotely by our tap recogniser.

interactions (e.g., [37]) have shown that callers would appreciate finer control of audio playback, so this was a natural application for our system. We were conscious of the need to explain usage of the service remotely, via an automated message played at the start of the voice site call, so chose to apply the technique to an interaction users were already familiar with.

In our implementation, when users tap two or three times on the back of their device, the time compression is 25 % and 35 %, respectively, while still retaining intelligibility. Tapping once returns playback to its normal speed. While this might seem an unusual choice of action for the gesture that is easiest to perform, it is worth bearing in mind that this is also potentially the most common gesture to be erroneously recognised. For this reason, we chose to ensure that false positives would have a non-intrusive effect on the flow of the call – if the user is listening at normal speed then false positives will have no effect at all.

Throughout the call, users are also still free to control the speed of playback by using traditional DTMF inputs instead of taps. In this case, keys 4, 5 and 6 correspond to single, double and triple taps, respectively.

The TapBack system is installed on a remote server, monitoring low-level network packets to track incoming phone calls to multiple voice sites. Figure 6.1 illustrates the operation of the system. When a call is established, real-time audio capturing, decoding and analysis is initialised. The incoming audio is first filtered to remove frequencies below 3 kHz, greatly reducing the problem of ambient noise. The stream is then windowed using a 512-sample Hamming window with an overlap of $\frac{7}{8}$. Tap recognition itself is very unsophisticated, simply searching each window for short, high-intensity, high-frequency sounds. Detected tap events are then fed to a higher-level audio gesture classifier which uses basic heuristics and gesture timeouts to classify each tap type.

When an audio gesture is found, the system creates a simulated DTMF packet and sends this as a command to the Spoken Web server, which adjusts the voice site playback speed in response to the request. By implementing the prototype in this way—as a network traffic monitor, rather than a system-specific component—it is potentially able to be used as an add-on to any interactive voice service, rather than just the Spoken Web. The voice service that receives the audio gesture can processes standard DTMF inputs normally, regardless of the input method that was actually used by the caller.

6.2.2 Experiment: testing remote tap recognition accuracy

We conducted an initial user study aiming to measure and improve the recogniser's accuracy over a standard telephone connection. Eighteen users of an existing, popular farming information voice site based in a rural region of Gujarat in India were recruited (see [109] for more detailed user population demographics). To ensure a cross-section of user expertise, participants included people who access the voice site very regularly and also those who are only casual users of the service. All users were male, and the average age was 32. The set of phones used by the participants consisted of 14 different (low-end) handset types produced by four manufacturers. All participants had already used the DTMF-based speed control methods that are currently available on the service (detailed in [37]).

Each participant was called by phone to explain the study method and the concept of audio gestures. The calls were made to participants when they were at locations from which they usually interact with the voice site services. The participant was then connected to a test voice site, which asked them to tap the back of the phone while holding it to their ear, in response to four sets of cues. Each cue set asked users to tap once; twice; then, three times. Each participant, therefore, provided 12 tap commands.

Gesture	Rate
1-tap	93 %
2-tap	78 %
3-tap	56 %

TABLE 6.1: Recognition rates of the three different tap-based audio gestures in our initial study.

Table 6.1 shows the recognition rates achieved during the study. High accuracy rates for single and double taps were encouraging given the external factors that were involved. Firstly, we gave only minimal explanation of the concept to users, and this form of interaction with a service was entirely novel in users' experience. A diverse set of low-end phones were involved, and the audio channel was of standard telephone quality, rather than the local microphone input commonly used in similar devices. Finally, the study was in a live, not laboratory setting – users participated while going about their day-to-day activities.

A large proportion of the errors in recognition were due to participants tapping slower than the recogniser expected. This led to 2-taps being recognised as $1 + 1$ taps (accounting for 50 % of the 2-tap errors) and 3-taps being recognised as $1 + 2$; $2 + 1$; or, $1 + 1 + 1$ taps (60 % of 3-tap errors). The remaining errors were caused by taps not being distinct enough for the recogniser to extract from the input.

6.3 Investigating rich back-of-device interactions in situ

We modified the prototype to improve its recognition performance based upon the results of our initial study, adding simple correction heuristics so that, for example, a 2-tap shortly followed by a 1-tap was interpreted as a 3-tap instruction. Following this, the system was made available on the farming information voice site.

This exploratory study aimed at measuring the adoption of audio gestures by logging any tap interactions and responses during normal use of the service. We also gathered more detailed feedback from a subset of participants in order to further understand the needs of the target users and the future potential for audio gestures.

6.3.1 Experiment: system deployment on a live voice site

The system was deployed on the voice site for 12 days, during which any of the 110 registered active users could call at any time. These users are geographically dispersed over a wide area of the state of Gujarat, and are all farmers living in rural settings. When calling, users were given a brief automated introduction to the new method that explained how they could tap the back of their phone to control the playback speed. The system logged call details and any input actions (both tap-based and DTMF, for comparison).

We supplemented these measures by conducting detailed telephone interviews with 15 users. Ten of these were selected at random from the set of those who had used the TapBack input method during the deployment period; the remaining five were randomly selected from people who called during this period but did not attempt to use the tap interaction. The average age of participants was 31, and all except one were male. During the interviews these users were asked about their reactions to the approach; how usable it was; any issues they had encountered in its use; and, suggestions for additional audio gestures and corresponding actions. Interviews were conducted in the participants' native language (Gujarati).

6.3.2 Findings

Table 6.2 shows the distribution of audio gestures recorded during the trial. A total of 286 calls to the voice site were recorded over the study period, from 52 unique callers. The TapBack feature was used by 36 people (166 calls), with 7.8 tap gestures per call, on average. Of the 36 participants that used tap interaction, 25 used the feature on more than one call. Two others called more than once but only used tap interaction on their first call; the remaining nine TapBack users called only once over the study period. Very few of the callers who wanted to control the speed of the

Gesture	Count
1-tap	772
2-tap	301
3-tap	220
Total	1293

TABLE 6.2: Audio gesture inputs detected during our second study.

call used the alternative DTMF-based method – 52 speed control DTMF events were recorded in total. The 16 participants who did not use tap interactions did not use the DTMF method for speed control, either.

The 1-tap gesture to return playback to normal speed was the most commonly recognised input. It is not possible to determine what proportion of these events were false positives recognised erroneously, however.[2] Of the double and triple tap events, similar proportions to those seen in the earlier recognition test were recognised as aggregates (e.g., double taps as two 1-taps, and triple taps as a sequence of 2- and/or 1-taps), but in this second trial these were recognised and corrected for.

Participant interviews

Considering first the ten participants who had used the TapBack feature during their calls. Of these, the majority were positive in their comments about the approach. Benefits mentioned ranged from those related to utility to those concerning the less-tangible 'user experience.' Several respondents commented on the tapping being easier to use and quicker than DTMF. Another interviewee talked of the 'fun' of the new interaction. Interestingly, one participant said, *"this is like having a touchscreen, this is a modern thing to use – it's cool."*

Negative comments from these ten adopters included the predictable, such as frustration when a tap-event was not recognised: one respondent said he would always use buttons, because *"they always work – end of story."* However, there were also issues related to the context of use. Two interviewees worried about using the system regularly as the tapping, to their mind, might damage the phone. For one of these interviewees this was particularly worrying as they often lent their phone to others to use (a practice quite common in rural low-income areas). Another respondent said they tended to listen to the service with a group of people using the speakerphone, so controlling the voice site via DTMF was less of an issue.

---

[2] More single taps were recognised than would be needed to return all calls to their normal speed after 2- and 3-tap gestures, but without examining each individual gesture manually it is not possible to determine whether these were intentional interactions or false positives.

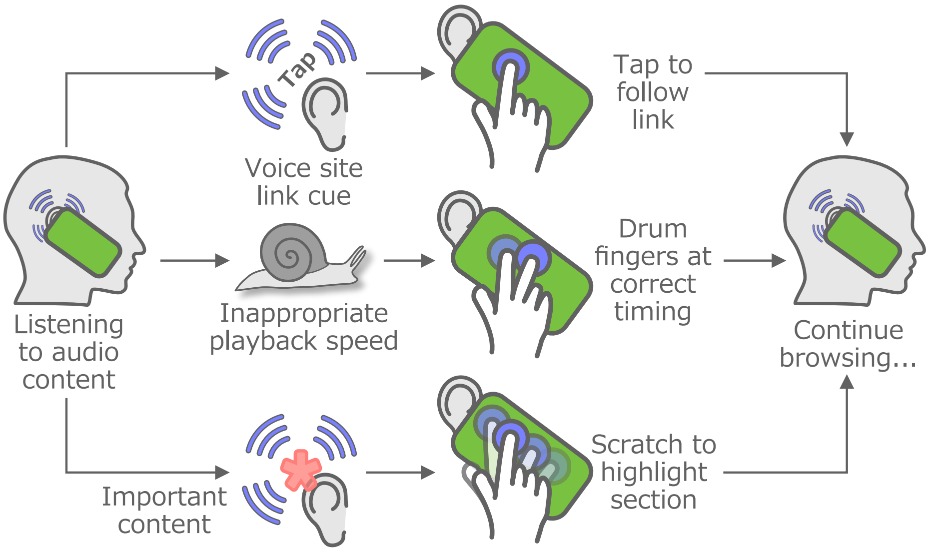


FIGURE 6.2: Potential future usage scenarios for audio gestures.

There were two explanations for the non-use of the approach by the five other interviewees. For some, their environments (as witnessed during telephone interviews) were too noisy. Others had not understood the new feature as explained by the voice site after the call was connected.

The majority of the 15 interviewees identified the value of two additional single-handed, back-of-phone interactions: drumming fingers and scratching. Many users suggested finger-clicking (using the non-phone holding hand). Most users also raised the possibility of making non-verbal utterances – for example: *"I could make the noise I make when shooing away cows."*

Interviewees found it hard to make mappings from their suggested gestures to voice site controls. Unlike many typical user study participants, who have extensive computer experience, these respondents had no notion of interface metaphors. However, one commented on the use of drumming to skip through voice content; and many others wanted a fast way to jump to particular sections of the audio. Figure 6.2 shows how we could widen the set of audio gestures in response to the participants' comments.

6.3.3 Discussion

The usage results from this study provide evidence that participants were willing to adopt the eyes-off tapping method, with the majority of callers using the approach. It should also be noted that the functions controlled by the tap prototype—speeding and slowing the audio—are optional: users are able to listen to and navigate through content without employing them. It should be expected, then, that some calls during the study would not show tap interaction. In light of this, the high adoption rate achieved is particularly encouraging. Furthermore, 93 % of callers who used the tapping input during their first call also used it again in their subsequent calls. Callers that used TapBack did so several times in each call.

The comments about the system's utility value are, of course, encouraging, especially when considering the relatively low accuracy of our simple recogniser. Earlier recognition issues appear to have been addressed in this second study, but further work is likely to be needed to improve its robustness for long-term usage. One potential approach could be to tailor tap recognition to each unique incoming phone number, adapting to the dynamics of the user's back-of-device surface and offering an improved recognition rate.

Participants in our interviews were, in most cases, keen to try a new interaction method, despite some underlying concerns about damaging their phones via tapping. Of more note, perhaps, are the responses relating to the user experience – a fun, 'modern' interaction is something that is not usually associated with the low-end devices that our participants have access to. We see these comments as spurs to further investigate the ways in which rich, eyes-off interaction methods can be provided for users who do not have the latest hardware.

The negative comments are an incentive to improve the recognition engine for future use. Our design currently uses a very simplistic method for tap parsing – we expect that recognition rates could be improved greatly with a refined model of tap interactions. It is worth considering negative feedback in light of the fact that users were not guided through a system demonstration; rather a short tutorial clip was played at the start of their first call. Despite this, a significant number of users adopted the technique.

The social issues that were raised suggest extensions to our approach. Currently we do not tailor tap recognition to individual users' inputs. Future versions will need to ensure that the tap models used are tuned not just to individual phone numbers but the set of users that might use that phone. For speaker-mode use, we might be able to consider a wider set of audio gestures as suggested by Harrison and Hudson

[54]. Alternatively, increased robustness might be achieved by simplifying the gesture set to allow only single taps. In the current speed control application, this could be applied as a toggle between options so each individual tap would change playback to the next speed preset.

6.4 Conclusions

In this chapter we have investigated the potential for offering advanced, eyes-off interactions to users of low-end devices. The use of remotely-recognised inputs has allowed us to complement current input methods and provide potentially eyes-off interactions regardless of the device that people are using.

Our prototype demonstrated how back-of-device tapping can be used as input by rural Indian farmers via their basic mobile handsets. These users' responses to the TapBack prototype, along with their suggestions for additional gestures, indicate the viability of the approach for people with very low exposure to computing. Furthermore, while the technique might appear unsophisticated or less exciting compared to the methods proposed for high-end mobiles, we argue that it is potentially far more likely to have impact in the sorts of contexts similar to those that we studied in this chapter than many of the higher-end methods we have investigated previously in this thesis.

6.4.1 Designing eyes-off interactions for impoverished platforms

While the system demonstrated in this chapter is clearly an early exploration of eyes-off interactions in developing contexts, the design and evaluation process has highlighted several benefits and recommendations for offering richer interactions for people without the latest devices.

It is clear from our results that remotely-recognised inputs can allow cheap, widely-available devices to use eyes-off interactions. The accuracy of such recognitions is likely to be highly dependent on the type of inputs chosen and the environment in which they are used, however. In our case the detection of taps was an obvious choice for a simple gesture recogniser; future extensions may need more complex recognition engines in order to be able to handle the various external factors that might affect accuracy. Offering rich inputs on standard devices can provide users with a more advanced interaction method than they might be familiar with, but it is important to be aware of situational concerns, such as those about device damage, that are perhaps less of an issue in more developed regions. The common practice of sharing devices—both between individuals, and while gathered in groups—should be taken into account when designing interactions. Furthermore, the difficulties users might have in conceptualising a new interaction style need to be carefully considered – users in our study struggled to understand interface metaphors, and so found it particularly difficult to offer feedback on potential usage.

Remote The use of remote input recognition can help designers provide rich interactions on low-end devices. Care should be taken when designing recognition engines, however – it is important to take into account the various (often unclear) external factors that could impact recognition rates and, consequently, the user experience.

Context Many people in developing regions are keen to use devices that are seen to be 'modern.' As previous research—and commercial failures—have illustrated, designing devices specifically for particular contexts is usually not the solution. However, providing rich, seemingly high-end interactions on the low-end phones that are common in these regions could potentially bridge this gap.

CHAPTER SEVEN

Conclusions

We began this thesis by lamenting the future of mobile interaction, arguing that it is misguided to spend such a significant amount of time looking at a screen to stay connected, located and up-to-date. Consequently, the designs and prototypes we have explored have aimed at providing more physically-grounded exploration of spaces – letting people point, feel or move to interact with their mobiles, eyes-off. A key goal throughout this research was to ground the everyday interactions we have with our devices in the physical world we live in, rather than the screen we use to view related digital content. However, we have not focused on completely removing the need for a screen. Instead, we have concentrated on re-imagining situations where techniques that are not screen-primary could complement existing methods. We have used alternative modalities—primarily haptic feedback—to show how people might interact directly with their surroundings, with the people around them, or just with their device, without the need to direct their attention toward the tool they are using.

Naturally, our aims changed and shifted throughout this work. At the start of this research we aimed to measure and compare the differences between eyes-off and screen-primary designs in both efficiency and time taken. In later research, however, our investigations moved to focus more on engagement with physical or digital spaces, and the user's immersion in their environment. We argue that this comparison gives a more accurate picture of people's usage of eyes-off designs, and also helps to better illustrate their suitability for the task at hand. These later experiments again used time as a measure, but compared participants' usage of our eyes-off designs to their normal behaviour, rather than an alternative system, in order to see whether engagement and cognitive load were affected.

Throughout this thesis we have argued for and demonstrated a class of designs that support this engaging eyes-off, physically-grounded interaction. Each of our four contributing chapters has added to the existing understanding and knowledge of mobile interaction, and the key findings from each chapter have been brought together into recommendations for future systems that explore this design space. In the next sections our core contributions are summarised and refined into concrete, concise guidelines for eyes-off, physically-grounded systems.

7.1 Summary of contributions and significant results

Throughout this thesis we have explored new ways of interacting with our mobiles, and with the environment around us. These techniques were embodied in design prototypes, each of which has been evaluated in one or more user studies.

In Chapter 1 we situated our work, considering the progress of 'invisible computing' from the original ubicomp designs to the present day. We argued that eyes-off—rather than simply eyes-free—interaction could complement, instead of replacing, existing techniques. In Chapter 2 we reviewed previous literature related to the areas we have subsequently explored in depth. We highlighted gaps in existing knowledge, and illustrated how previous work has not fully exploited the possibilities available for eyes-off, physically-grounded and mobile location-aware interaction.

Chapter 3 demonstrated the benefits that can be found in less-screen-focused designs for pointing-based discovery of geolocated content. We began by investigating three progressively less-visual systems, showing that while screen-based feedback allows for more accuracy, the less visual versions support faster and more exploratory interaction. Participants appreciated the low effort required to discover targets using the eyes-off design.

In the second half of this chapter we turned to explore the augmentation of a pointing-based discovery system with haptic feedback. Participants in a user study were able to use the system while simultaneously walking and navigating around obstacles. The use of vibrotactile feedback allowed greater accuracy in the pointing interaction and, in two-thirds of cases, was not significantly different to a visual comparison system. Additionally the use of feedback allowed participants to select content hotspots with fewer false positives than the screen-primary alternative.

Chapter 4 looked deeper into the process of in situ content retrieval, investigating three separate designs for displaying geolocated elements. Our first design investigated the use of handheld projectors, but this was ultimately rejected in favour of multi-level haptic feedback, which allowed users to discover the presence of different content categories without the need to look at a screen. Our two studies of this system demonstrated both its usability and its accuracy, again highlighting participants' preference for exploratory interaction rather than direct and immediate display of discovered content.

Our third prototype in this chapter turned to consider how the locations of other people, rather than geolocated content, could be displayed in an eyes-off manner. This collaborative project demonstrated how navigation through a space could be a function of the locations of other people in the same area. Our haptic meetup prototype allowed groups of people to display the location of, and ultimately navigate to, a mutually-convenient central rendezvous point. The system was extensively simulated prior to the trial, significantly reducing the time required for testing and configuration, and allowing selection of the most appropriate interaction parameters. As a result, participants were easily able to find and follow the wide-angle vibration zone, and could use this intelligently to choose an appropriate route to the meeting point, with minimal cost in performance when compared to navigation to a known goal.

Chapter 5 continued the theme of eyes-off navigation, looking at several methods for shaping people's understanding of and movement through a physical space. Our initial design showed how fixed-width haptic feedback can support and enable pedestrian navigation over large distances. This was extended in, and compared to, our second prototype, which offered variable-width haptic feedback as a function of the number of potential routes that were available from the user's position. Participants in our large-area user study were empowered to choose their own paths, but confident that the underlying directional feedback would help them find their goal. The system demonstrated how the common implementation of GPS navigation—with constant turn-by-turn instructions—may not be particularly relevant to pedestrian navigation, where autonomy can be beneficial in many cases.

In the second half of this chapter we extended the variable-width vibrotactile feedback concept to show how existing social media content from people outside of the immediate area might be used to shape a user's navigational behaviour. We simulated three potential designs for social navigation, and demonstrated their potential for helping prompt exploration of particularly interesting or popular areas. The results from our simulations hint at the potential for automatic off-the-beaten-track tours of both urban and rural areas, depending on the social content source that is selected.

In Chapter 6 we turned to investigate how the eyes-off interactions we explored in Chapters 3 to 5 might be made more inclusive by offering them on devices that are not able to support the necessary positioning, sensory and computational requirements. We created a prototype audio gesture recogniser which enabled back-of-device tap inputs on any phone, and deployed the system on a voice-based information service used by farmers in rural India. Our studies of the system demonstrated both its technical capability and its adoption by regular users of the service. Comments from participants about this eyes-off input method showed how it was thought of as a modern and 'cool' service – reactions that are not typically common when discussing low-end phones.

7.2 Design recommendations for eyes-off physically grounded interaction

Throughout this thesis we have offered contributions to the understanding and design of future eyes-off and physically-grounded devices. At the conclusion of this work these may now be distilled into four key design recommendation themes for future systems that explore this area.

Pointing Pointing as an interaction method can be accurate, but the gestures used should be as straightforward as possible. Multi-level and multi-gesture pointing interactions are likely to be less accurate in both components of the gesture.

Point-to-select gestures are most accurate at short distances, due to minor errors becoming magnified when users attempt to point to objects further away. It is difficult for users to convert observed physical distances into a system input, and accuracy is likely to decrease as the target distance increases.

Feedback To improve the accuracy of eyes-off pointing gestures, directional feedback should be used where possible. Without feedback, eyes-off designs can offer only broad measures of accuracy; with feedback, precision is greatly improved.

The size of feedback zones, and the delivery of feedback within these, have a large impact on the user experience. Haptic target zones should be sizeable enough for users to discover them easily, but not so large that performance suffers if users engage in edge-following behaviour. For navigation applications, 60° is ideal; for target selection the size used will depend on the number of targets that need to be represented.

Pedestrians are able to track the position of a moving vibrotactile target eyes-off, but find it difficult to feel steadily increasing or decreasing changes in intensity. As a result, intensity differences between feedback zones are likely not necessary, but if employed the boundaries should be clearly distinct, rather than gradual.

Exploration Vibrotactile eyes-off feedback should be used primarily to support exploratory interaction – whether to explore a space as a pedestrian, or to browse the digital content that augments the location.

Low-resolution directional vibrotactile feedback can help pedestrians navigate to a goal without the need to look at the device that is guiding them. Designers of future systems in this area should aim to maintain this eyes-off interaction method where possible, but fall back to more traditional navigation designs where necessary – at major path junctions, for example, or towards the end of the navigation process.

Offering feedback that varies in width can help empower pedestrians to make their own route choices through a space. People are able to fuse guidance offered by directional vibrotactile feedback with their own initiative and instincts.

The need for precise, instructive and controlling turn-by-turn waypoints is far less evident for pedestrian navigation than it is for traditional, car-based GPS navigation. Waypoint-free designs are usable over distances of at least 1.5 km. The use of this technique for typical city or urban navigation scenarios, in which the distances are relatively small, seems likely to be beneficial.

Development It is common for people in developing regions to be keen to use devices that are seen to be 'modern.' Designing devices specifically for particular contexts, or handing out expensive hardware, is likely not the solution. Instead, the use of remote input recognition—particularly over a voice-based information service such as the Spoken Web—can help provide rich interactions on the low-end devices that are common in many developing areas.

7.3 Limitations and generalisability

The research presented in this thesis has arguably already offered lasting contributions to the designs of future devices – its impact has been recognised by international researchers in the form of citations, by media organisations through news articles, and finally in the adoption of our remote audio gesture recognition design on the Spoken Web platform. It is important, however, to highlight the limitations of this work.

The prototypes we have developed and demonstrated throughout this thesis have been tested and evaluated in user studies, but we have not investigated the use of our systems in longitudinal trials. A longer-term trial of the interaction designs proposed in this thesis would help to highlight the ways in which users appropriate the devices for their own purposes, and remove any effects in our studies that have been caused by the novelty of the prototypes. Additionally, the devices and services we have created have each been standalone – while there is a demonstrated desire by users to adopt many of the interaction styles we have used, we have not been able to test, longer-term, a single device or application that embodies all of these features.

We used Swansea University campus as a study location for many of the outdoor studies undertaken for this research. This particular campus is ideal for outdoor studies, as it contains many features found in both urban and rural locations – from compact, tightly-spaced building and pathway sections, to more open built spaces, through to rolling parkland and wooded areas. Results from our measures of cognitive load and walking speed (PPWS) showed little difference in user behaviour between these types of environment, suggesting that our results are generalisable to at least these cases. There are of course limitations to this result, however, such as potential user familiarity with the study area and the geographical size of the campus locations studied. Further trials are necessary to fully validate our systems in a wider range of environments.

Our results from simulations are, we argue, generalisable to similar physical environments. In each of our simulated trials, 500 iterations were run, and the results were well matched to actual results from field trials. While the simulator used was simplistic with regards to human wayfinding and path choice reasoning, its models of variation in walking rate, interaction time, network delay, and environmental influences, such as GPS noise and update rate, helped to make it a robust and accurate model of the quantitative results of a real trial. Simulations are not able to reveal qualitative data, however, such as user opinions and behaviour while using a system. For this reason our simulations are limited to providing estimates of timings and general navigational results – further work is needed to gather user feedback on our simulated designs.

Many of the studies in this thesis were undertaken using between-groups methods, or consisted of only a single system. The main reason for this is primarily operational – the complexity of planning, coordinating and running multiple-participant, large-area trials is significant. For example, recruiting, positioning, tracking and managing each set of participants over the large area used for our group navigation prototype was a highly time consuming—and costly—task, requiring at least one study assistant per participant to help manage the process. Future work could replicate this type of trial in additional scenarios, and with varying numbers of participants, in order to help reproduce and broaden the scope of our results.

7.4 Future work

There are several additional studies and variations that could be undertaken to further extend the results in this thesis. For instance, in Chapter 4 the walking speed we chose to simulate was relatively slow when compared to the actual speed at which participants walked – further work could examine additional variations in walking speed (3 km/h to 7 km/h, for example) to illustrate the affects that these have on task duration. Further possibilities are present in, for example, our dynamic navigation prototype, which could be extended to deal with cases where a single available path crosses an insurmountable barrier (e.g., a river or highway). In this case, there is a clear need to switch from pure goal-based navigation to some sort of directional cue. Future versions of this type of navigation system will likely follow the chaperone paradigm defined by Graham and Cheverst [51], in which the system only interacts with the user if they are doing something wrong. This could be achieved either by repositioning the tactile zone to offer precise directions or, perhaps more appropriately, adding an additional directional hint (e.g., via [86]), using tactile effects to produce a distinctive indication that a specific path must be taken.

We have seen in the earlier chapters of this work how users often prefer exploratory browsing and displaying of geotagged content rather than immediate retrieval. This temporal disparity between selection and perusal—while common in the form of internet or physical bookmarks, for example—has not been widely investigated for physically-grounded interaction, and mainstream applications of this technique are lacking. A single application combining the key features of each of our discovery and displaying prototypes could support both exploratory discovery and delayed browsing.

More work is clearly needed to improve and clarify the tactile cues that we used for eyes-off content discovery. While the wider and variable vibrotactile zones we used in later chapters are demonstrably successful for navigation, it is unlikely that these will be sufficient or appropriate for haptic browsing of tightly-spaced content hotspots. Similarly, our early prototype using a linear model of point-and-tilt interaction is likely not a good match for how people mentally model distances. Interdisciplinary work linking gestural input to models of human cognition could help clarify how this interaction might be improved.

The social possibilities for the types of dynamic, fused interaction spaces we explored in our navigation prototypes are substantial, and the ability of users to engage via such spaces will grow as further interaction techniques are developed. Future extensions of this work could look more closely at how these social 'maps' might be created, evaluating on a larger scale using realtime social data. Extending this further, the incorporation of social or route data into a publicly-released prototype could form the basis for a class of navigation device offering users a choice between the most appropriate navigation methods, supporting exploration if desirable, or presenting waypoint-based shortest-path navigation if necessary.

Throughout this thesis we have avoided proposing the use of vibrotactile feedback for visually impaired users, recognising that when sight is poor, system errors and inaccuracies will have a much greater impact, and the more familiar, more sensitive, and much more robust current methods are far more appropriate. However, with the success of the tactile navigation demonstrated in the later chapters of this thesis, refined in line with the findings of previous work in this area (e.g., [12]), it is possible that navigation for visually impaired or blind users is now a viable complement to help spur exploration of physical spaces where suitable.

Future work in this area may first need to investigate the social implications of the eyes-off pointing gestures we have used in many of our prototypes. We have assumed throughout this work that, in a similar way to how the once-strange practice of walking down a street talking into a Bluetooth headset has become common, in situ pointing to people, places, or even to nothing in particular, might become socially acceptable. Further work is necessary, however, in order to validate this assumption.

Finally, there is much work still to be done to improve the performance of our remote audio gesture recogniser. Future work could focus on refinements to the recognition algorithms to improve accuracy by personalising the gestures or, alternatively, on simplifying the gesture set to minimise ambiguity.

7.5 Concluding remarks

When this research was started, in early 2008, high-quality touch screens were only just emerging, mobile phone-based GPS access was sparse and unreliable, and few devices contained the inertial sensors that we have employed throughout our work – even if sensors were present they were rarely used for anything other than basic screen rotation. Now, in 2012, even everyday mobile devices have a full-size capacitive touch screen, constant location awareness and internet connectivity, and a full complement of embedded sensors. As a result, the majority of the prototypes we have explored in this work can now be implemented as a downloadable application for any modern smartphone. Specialised hardware is no-longer necessary, and the barriers to entry for this type of mobile development have been almost completely removed.

This incredible transition should be seen as an exciting incentive to further explore the types of interaction we have demonstrated in this thesis. While it is now possible for anyone to create an eyes-off, physically-grounded mobile application, a casual browsing of the smartphone application markets shows that there is much work yet to be done. A reading of recent publications in this area (e.g., [72, 150, 153, 163]) suggests that eyes-off design, at least as a concept, is steadily increasing in popularity with interaction researchers. Still, we argue that the most fruitful areas for research impact are, as we began to explore for our final chapter, in developing methods and techniques for enabling this class of interactions for the millions of people who do not currently have the means to access them.

APPENDIX A

Contributing Publications

Much of the research within this thesis has previously been published in international peer-reviewed conference or journal papers. Abstracts of the nine publications that are a major part of this work are reproduced in this section for ease of reference. Where significant parts of the work have been collaborative, a summary of my contribution is also included. I was the main author of all publications except one (see individual summaries, below). All publications except [P1] were co-authored by Matt Jones, whose contribution to this work is immeasurable – like most PhD theses this research has in many places been a collaboration, and as such fuses many ideas and diverse viewpoints.

The impact of this research has been recognised in several forums – particularly, [P5] was the winner of the 'best paper' award at MobileHCI 2009, and [P7] was nominated for the same award in the following year. Publication [P7] was also recognised outside academia, being featured in New Scientist magazine,[1] and on CBC Radio-Canada,[2] amongst other media venues. A patent application was submitted by IBM research for the concepts described in [P9]. The impact of the work has also been noted by other researchers in the field, with many receiving multiple citations.

Publications are listed below in the order in which elements from them have appeared in this thesis.

---

[1] *Navigation app gives you freedom to explore* – see: http://www.newscientist.com/article/mg20727775.800-navigation-app-gives-you-freedom-to-explore.html

[2] *As It Happens* – see: http://www.cbc.ca/video/news/audioplayer.html?clipid=1593970444

[P1] Robinson, S. Heads-up engagement with the real world: multimodal techniques for bridging the physical-digital divide. In *Proc. CHI '10: Extended Abstracts*, ACM (2010), 2895–2898.

Abstract The vast and ever-increasing collection of geo-tagged digital content about the physical world around us has prompted the development of interaction methods for various different scenarios. However, the map-based views common on desktop computers are not always appropriate when considering mobile usage. The aim of this research is to provide suitable methods that can encourage user interaction with geo-located digital content, avoiding unnecessary interference with the user's immersion in the physical world around them. This extended abstract outlines the work published to date, suggests future areas of research, and highlights the key contributions brought to the HCI community.

Author's contribution This Doctoral Consortium extended abstract, and the CHI 2010 DC session I participated in as a result, helped to bring together and formulate many of the ideas discussed throughout this thesis. The concepts and writing of the paper were mine, with feedback and advice from colleagues and supervisors.

[P2] Robinson, S., Eslambolchilar, P. and Jones, M. Exploring casual point-and-tilt interactions for mobile geo-blogging. *Personal and Ubiquitous Computing* 14.4 (2010), 363–379.

Abstract People record and share their experiences through text, audio and video. Increasingly they do this blogging from mobile devices. We illustrate a novel, mobile, low interaction cost approach to supporting the creation of a rich record of journeys made and places encountered. By pointing and tilting a mobile, users indicate their interests in a location. No content is provided to the user in situ but, later, web materials including images, entries from other people's blogs and web pages are automatically placed on an interactive map for viewing on a larger screen device. We built two mobile prototypes to explore the approach – one combines gestures and visual map feedback; the other is more lightweight, allowing the user to simply point-and-tilt. We describe and motivate the approaches and present user studies that raise issues relevant to their design and to the wider class of device and service concerned with mobile spatial information access.

Author's contribution Portions of the first section of this article were previously submitted as part of an MSc dissertation; those sections are not included in this thesis. The design, implementation and analysis behind the work was mine. I wrote the journal article, with feedback and advice from collaborators and other authors.

[P3] Robinson, S., Eslambolchilar, P. and Jones, M. Evaluating haptics for information discovery while walking. In *Proc. BCS-HCI '09*, British Computer Society, (2009), 93–102.

Abstract In this article we describe and evaluate a novel, low interaction cost approach to supporting the spontaneous discovery of geo-tagged information while on the move. Our mobile haptic prototype helps users to explore their environment by providing directional vibrotactile feedback based on the presence of location data. We conducted a study to investigate whether users can find these targets while walking, comparing their performance when using only haptic feedback to that when using an equivalent visual system. The results are encouraging, and here we present our findings, discussing their significance and issues relevant to the design of future systems that combine haptics with location awareness.

Author's contribution The concept and implementation behind this research was mine. I planned and ran the study sessions, analysed the results and wrote the paper with feedback from other authors.

[P4] Robinson, S. and Jones, M. HaptiProjection: multimodal mobile information discovery. In *Proc. Ubiprojection Workshop at Pervasive '10*, 2010.

Abstract Handheld projectors are steadily emerging as a potential display method of the future, offering many opportunities for interesting interactions with the world around us. However, to date little attention has been focused how people might move from mobile device usage to projection of interactive content. In this position paper we address this by proposing a method for location-based content discovery that helps to merge the physical and digital spaces we live in. We describe an early prototype, developed to demonstrate interaction concepts, and summarise the challenges and future developments needed for this type of system.

Author's contribution The concept for this workshop paper was mine – I developed the system and wrote the paper, with feedback from Matt Jones.

[P5] Robinson, S., Eslambolchilar, P. and Jones, M. Sweep-Shake: finding digital resources in physical environments. In *Proc. MobileHCI '09*, ACM (2009), 85–94. **Winner of *best paper*.**

Abstract In this article we describe the Sweep-Shake system, a novel, low interaction cost approach to supporting the spontaneous discovery of geo-located information. By sweeping a mobile device around their environment, users browse for interesting information related to points of interest. We built a mobile haptic prototype which encourages the user to explore their surroundings to search for location information, helping them discover this by providing directional vibrotactile feedback. Once potential targets are selected, the interaction is extended to offer an hierarchy of information levels with a simple method for filtering and selecting desired types of data for each geo-tagged location. We describe and motivate our approach and present a short field trial to situate our design in a real environment, followed by a more detailed user study that compares it against an equivalent visual-based system.

Author's contribution Four of the study sessions described in this paper were overseen by Parisa Eslambolchilar. The design and development of the system were mine, along with the remainder of the study sessions, analysis and paper authoring.

[P6] Williamson, J., Robinson, S., Stewart, C., Murray-Smith, R., Jones, M. and Brewster, S. Social gravity: a virtual elastic tether for casual, privacy-preserving pedestrian rendezvous. In *Proc. CHI '10*, ACM (2010), 1485–1494.

Abstract We describe a virtual "tether" for mobile devices that allows groups to have quick, simple and privacy- preserving meetups. Our design provides cues which allow dynamic coordination of rendezvous without revealing users' positions. Using accelerometers and magnetometers, combined with GPS positioning and non-visual feedback, users can probe and sense a dynamic virtual object representing the nearest meeting point. The Social Gravity system makes social bonds tangible in a virtual world which is geographically grounded, using haptic feedback to help users rendezvous. We show dynamic navigation using this physical model-based system to be efficient and robust in significant field trials, even in the presence of low-quality positioning. The use of simulators to build models of mobile geolocated systems for pre-validation purposes is discussed, and results compared with those from our trials. Our results show interesting behaviours in the social coordination task, which lead to guidelines for geosocial interaction design. The Social Gravity system proved to be very successful in allowing groups to rendezvous efficiently and simply and can be implemented using only commercially available hardware.

Author's contribution The majority (approximately 90 %) of the system simulation and prototype development for this research was by John Williamson, with portions of this by Craig Stewart. John also contributed the extensive walking speed analysis that is discussed in the paper (but not in this thesis), building upon previous work by Andrew Crossan [33]. I refined the prototype, and undertook the study design, planning and management, and the analysis that is discussed in this thesis. Study sessions were run with the help of Richard Byrne, Chris Elsmore, Darius Garnham, Fernando Loizides, Patrick Oladimeji, Tom Owen and Jennifer Pearson. Authorship of the paper was split approximately evenly between the six authors. Only the sections I wrote are included in this thesis.

[P7] Robinson, S., Jones, M., Eslambolchilar, P., Murray-Smith, R. and Lindborg, M. "I did it my way": moving away from the tyranny of turn-by-turn pedestrian navigation. In *Proc. MobileHCI '10*, ACM (2010), 341–344. **Nominated for *best paper*.**

Abstract In this article we describe a novel approach to pedestrian navigation using haptic feedback. People are guided in the general direction of their destination via vibration, but additional exploratory navigation is stimulated by varying feedback based on the potential for taking alternative routes. We describe two mobile prototypes that were created to examine the possible benefits of the approach. The successful usage of this exploratory navigation method is demonstrated in a realistic field trial, and we discuss the promising results and interesting participant behaviours that were recorded.

Author's contribution The simulations described in this paper were adapted from the simulator built by John Williamson for [P6]. The remainder of the design, development, study, analysis and paper authoring were mine.

[P8] Robinson, S., Jones, M., Williamson, J., Murray-Smith, R., Eslambolchilar, P. and Lindborg, M. Navigation your way: from spontaneous independent exploration to dynamic social journeys. *Personal and Ubiquitous Computing* 16.8 (2012), 973–985.

Abstract In this article, we describe a novel approach to pedestrian navigation using bearing-based haptic feedback. People are guided in the general direction of their destination via a minimal directional cue, but additional exploration is stimulated by varying feedback based on the potential for taking alternative routes. This extreme navigation method removes the complexities of maps and direction following, concentrating on allowing pedestrians to actively explore their surroundings, rather than offering perfect, but passive, turn-by-turn guidance. We simulate and build two mobile prototypes to examine the possible benefits of this approach, then further extend its impact by considering how social media might be incorporated to provide a real-time, dynamically evolving map of physical locations. The successful use of our mobile prototypes is demonstrated in a realistic field trial, and we discuss the results and interesting participant behaviours that were recorded, validating the predictions from their earlier simulation. We continue by simulating the use of publicly posted status updates and pictures as a proxy for location mapping, showing how these methods can produce comparable navigation results to real-world field trials, highlighting their potential as tools for real-world social journeys.

Author's contribution This research used the simulator from [P6] as a base for building the three social media navigation prototypes that are discussed and evaluated. Portions of the paper's text were extended by Roderick Murray-Smith, and John Williamson contributed the description of the simulator's design and rationale; these are not included in this thesis. The design, development, study, analysis and remainder of the paper authoring were mine, with feedback and advice from other authors.

[P9] Robinson, S., Rajput, N., Jones, M., Jain, A., Sahay, S. and Nanavati, A. TapBack: towards richer mobile interfaces in impoverished contexts. In *Proc. CHI '11*, ACM (2011), 2733–2736.

Abstract Much of the mobile work by HCI researchers explores a future world populated by high-end devices and relatively affluent users. This paper turns to consider the hundreds of millions of people for whom such sophistication will not be realised for many years to come. In developing world contexts, people will continue to rely on voice-primary interactions due to both literacy and economic reasons. Here, we motivate research into how to accommodate advanced mobile interface techniques while overcoming the handset, data-connection and user limitations. As a first step we introduce TapBack: back-of-device taps to control a dialled-up, telephone-network-based voice service. We show how these audio gestures might be recognised over a standard telephone connection, via users' existing low-end devices. Further, in a longitudinal deployment, the techniques were made available on a live voice service used by rural Indian farmers. Results from the study illustrate the desire by users to adopt the approach and its potential extensions.

Author's contribution The design and development of the prototype discussed in this paper, design and analysis of the results of the user studies undertaken, and paper authoring, were mine. The study sessions were conducted at IBM Research India (New Delhi) by Anupam Jain, Amit Nanavati, Nitendra Rajput and Shrey Sahay.

Bibliography

[1] Ahmaniemi, T., Lantz, V. and Marila, J. Dynamic audiotactile feedback in gesture interaction. In *Proc. MobileHCI '08*, ACM (2008), 339–342.

[2] Aretz, A. The design of electronic map displays. *Human Factors* 33.1 (1991), 85–101.

[3] Aretz, A. and Wickens, C. The mental rotation of map displays. *Human Performance* 5.4 (1992), 303.

[4] Aslan, I., Schwalm, M., Baus, J., Krüger, A. and Schwartz, T. Acquisition of spatial knowledge in location aware mobile pedestrian navigation systems. In *Proc. MobileHCI '06*, ACM (2006), 105–108.

[5] Axup, J., Viller, S. and Bidwell, N. Usability of a mobile, group communication prototype while rendezvousing. In *Proc. CTS '05*, IEEE (2005), 24–31.

[6] Barnard, L., Yi, J., Jacko, J. and Sears, A. Capturing the effects of context on human performance in mobile computing systems. *Personal and Ubiquitous Computing* 11.2 (2007), 81–96.

[7] Beardsley, P., Baar, J. van, Raskar, R. and Forlines, C. Interaction using a handheld projector. *Computer Graphics and Applications* 25.1 (2005), 39–43.

[8] Bidwell, N. and Lueg, C. Creating a framework for situated way-finding research. In *Proc. APCHI '04*, LNCS 3101. Springer Berlin / Heidelberg (2004), 40–49.

[9] Bolt, R. ''Put-that-there'': voice and gesture at the graphics interface. In *Proc. SIGGRAPH '80*, ACM (1980), 262–270.

[10] Bornträger, C., Cheverst, K., Davies, N., Dix, A., Friday, A. and Seitz, J. Experiments with multi-modal interfaces in a context-aware city guide. In *Proc. MobileHCI '03*, LNCS 2795. Springer Berlin / Heidelberg (2003), 116–130.

[11] Bosman, S., Groenendaal, B., Findlater, J., Visser, T., Graaf, M. de and Markopoulos, P. GentleGuide: an exploration of haptic output for indoors pedestrian guidance. In *Proc. MobileHCI '03*, LNCS 2795. Springer Berlin / Heidelberg (2003), 358–362.

[12] Bradley, N. and Dunlop, M. An experimental investigation into wayfinding directions for visually impaired people. *Personal and Ubiquitous Computing* 9.6 (2005), 395–403.

[13] Brewster, S. Head-up interaction: can we break our addiction to the screen and keyboard? In *Proc. ICMI-MLMI '09*, ACM (2009), 151–152.

[14] Brewster, S. and Hughes, M. Pressure-based text entry for mobile devices. In *Proc. MobileHCI '09*, ACM (2009), 73–76.

[15] Brewster, S., Lumsden, J., Bell, M., Hall, M. and Tasker, S. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proc. CHI '03*, ACM (2003), 473–480.

[16] Brewster, S., Räty, V.-P. and Kortekangas, A. Earcons as a method of providing navigational cues in a menu hierarchy. In *Proc. BCS-HCI '96*, Springer-Verlag (1996), 169–183.

[17] Brown, B., Sellen, A. and O'Hara, K. A diary study of information capture in working life. In *Proc. CHI '00*, ACM (2000), 438–445.

[18] Brown, L., Brewster, S. and Purchase, H. Multidimensional tactons for non-visual information presentation in mobile devices. In *Proc. MobileHCI '06*, ACM (2006), 231–238.

[19] Burnett, G. Designing and evaluating in-car user-interfaces. *Handbook of Research on User-Interface Design and Evaluation for Mobile Technology* (2008), 218–236.

[20] Burnett, G. and Lee, K. The effect of vehicle navigation systems on the formation of cognitive maps. In *Proc. ICTTP '04*, Elsevier (2005), 407–418.

[21] Central Intelligence Agency. The world factbook: country comparison: telephones – mobile cellular. 2011. URL: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2151rank.html (visited on 01/12/2011).

[22] Chang, A. and O'Sullivan, C. Audio-haptic feedback in mobile phones. In *Proc. CHI '05: Extended Abstracts*, ACM (2005), 1264–1267.

[23] Cheverst, K., Davies, N., Mitchell, K. and Efstratiou, C. Using context as a crystal ball: rewards and pitfalls. *Personal and Ubiquitous Computing* 5.1 (2001), 8–11.

[24] Cheverst, K., Davies, N., Mitchell, K. and Friday, A. Experiences of developing and deploying a context-aware tourist guide: the GUIDE project. In *Proc. MobiCom '00*, ACM (2000), 20–31.

[25] Cheverst, K., Davies, N., Mitchell, K., Friday, A. and Efstratiou, C. Developing a context-aware electronic tourist guide: some issues and experiences. In *Proc. CHI '00*, ACM (2000), 17–24.

[26] Cheverst, K., Mitchell, K. and Davies, N. Investigating context-aware information push vs. information pull to tourists. In *Proc. MobileHCI '01*, 2001.

[27] Cheverst, K., Mitchell, K. and Davies, N. Exploring context-aware information push. *Personal and Ubiquitous Computing* 6.4 (2002), 276–281.

[28] Church, K., Neumann, J., Cherubini, M. and Oliver, N. SocialSearchBrowser: a novel mobile search and information discovery tool. In *Proc. IUI '10*, ACM (2010), 101–110.

[29] Colbert, M. A diary study of rendezvousing: implications for position-aware computing and communications for the general public. In *Proc. GROUP '01*, ACM (2001), 15–23.

[30] Colbert, M. Rendezvousing at familiar and unfamiliar places. *Journal of Navigation* 57.3 (2004), 327–338.

[31] Costanza, E., Inverso, S., Allen, R. and Maes, P. Intimate interfaces in action: assessing the usability and subtlety of EMG-based motionless gestures. In *Proc. CHI '07*, ACM (2007), 819–828.

[32] Cowan, L. and Li, K. ShadowPuppets: supporting collocated interaction with mobile projector phones using hand shadows. In *Proc. CHI '11*, ACM (2011), 2707–2716.

[33] Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J. and Musizza, B. Gait phase effects in mobile interaction. In *Proc. CHI '05*, ACM (2005), 1312–1315.

[34] Crossan, A., Williamson, J., Brewster, S. and Murray-Smith, R. Wrist rotation for interaction in mobile contexts. In *Proc. MobileHCI '08*, ACM (2008), 435–438.

[35] Dearman, D., Hawkey, K. and Inkpen, K. Rendezvousing with location-aware devices: enhancing social coordination. *Interacting with Computers* 17.5 (2005), 542–566.

[36] Dearman, D., Inkpen, K. and Truong, K. Mobile map interactions during a rendezvous: exploring the implications of automation. *Personal and Ubiquitous Computing* 14.1 (2010), 1–13.

[37] Dhanesha, K., Rajput, N. and Srivastava, K. User driven audio content navigation for spoken web. In *Proc. MM '10*, ACM (2010), 1071–1074.

[38] Dourish, P. Where the Action Is: The Foundations of Embodied Interaction. MIT Press (2001).

[39] Dourish, P. What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8.1 (2004), 19–30.

[40] Dourish, P. and Chalmers, M. Running out of space: models of information navigation. In *Proc. HCI '94*, BCS (1994).

[41] Egenhofer, M. Spatial information appliances: a next generation of geographic information systems. In *Proc. GeoInfo '99*, 1999.

[42] Endsley, M. Toward a theory of situation awareness in dynamic systems. *Human Factors* 37.1 (1995), 32–64.

[43] Erp, J. van. Vibrotactile spatial acuity on the torso: effects of location and timing parameters. In *Proc. WHC '05*, IEEE (2005), 80–85.

[44] Erp, J. van, Veen, H. van, Jansen, C. and Dobbins, T. Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception* 2.2 (2005), 106–117.

[45] Espinoza, F., Persson, P., Sandin, A., Nyström, H., Cacciatore, E. and Bylund, M. GeoNotes: social and navigational aspects of location-based information systems. In *Proc. Ubicomp '01*, LNCS 2201. Springer Berlin / Heidelberg (2001), 2–17.

[46] Floyd, R. Algorithm 97: shortest path. *Communications of the ACM* 5.6 (1962), 345–345.

[47] Fröhlich, P., Oulasvirta, A., Baldauf, M. and Nurminen, A. On the move, wirelessly connected to the world. *Communications of the ACM* 54.1 (2011), 132–138.

[48] Fröhlich, P., Simon, R., Baillie, L. and Anegg, H. Comparing conceptual designs for mobile access to geo-spatial information. In *Proc. MobileHCI '06*, ACM (2006), 109–112.

[49] Geelhoed, E., Falahee, M. and Latham, K. Safety and comfort of eyeglass displays. In *Proc. HUC '00*, Springer-Verlag (2000), 236–247.

[50] Goodman, J., Gray, P., Khammampad, K. and Brewster, S. Using landmarks to support older people in navigation. In *Proc. MobileHCI '04*, LNCS 3160. Springer Berlin / Heidelberg (2004), 37–57.

[51] Graham, C. and Cheverst, K. Guides, locals, chaperones, buddies and captains: managing trust through interaction paradigms. In *MobileHCI '04 workshop: HCI on Mobile Guides*, 2004.

[52] Guinard, D., Streng, S. and Gellersen, H. Extending mobile devices with spatially arranged gateways to pervasive services. In *Pervasive '07 Workshop: Pervasive Mobile Interaction Devices*, 2007.

[53] Hallnäs, L. and Redström, J. Slow technology – designing for reflection. *Personal and Ubiquitous Computing* 5.3 (2001), 201–212.

[54] Harrison, C. and Hudson, S. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proc. UIST '08*, ACM (2008), 205–208.

[55] Harrison, S. and Dourish, P. Re-place-ing space: the roles of place and space in collaborative systems. In *Proc. CSCW '96*, ACM (1996), 67–76.

[56] Hart, S. and Staveland, L. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Human Mental Workload* 1 (1988), 139–183.

[57] Heidegger, M. Being and Time. Wiley-Blackwell (1962).

[58] Hemmert, F., Hamann, S., Löwe, M., Zeipelt, J. and Joost, G. Weight-shifting mobiles: two-dimensional gravitational displays in mobile phones. In *Proc. CHI '10: Extended Abstracts*, ACM (2010), 3087–3092.

[59] Hemmert, F., Joost, G., Knörig, A. and Wettach, R. Dynamic knobs: shape change as a means of interaction on a mobile phone. In *Proc. CHI '08: Extended Abstracts*, ACM (2008), 2309–2314.

[60] Hinckley, K., Pierce, J., Sinclair, M. and Horvitz, E. Sensing techniques for mobile interaction. In *Proc. UIST '00*, ACM (2000), 91–100.

[61] Hoggan, E., Crossan, A., Brewster, S. and Kaaresoja, T. Audio or tactile feedback: which modality when? In *Proc. CHI '09*, ACM (2009), 2253–2256.

[62] Holland, S., Morse, D. and Gedenryd, H. AudioGPS: spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing* 6.4 (2002), 253–259.

[63] Hornecker, E. and Buur, J. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proc. CHI '06*, ACM (2006), 437–446.

[64] Hughes, S. SHAKE SK6 and SK7 Information Pages. 2012. URL: `http://code.google.com/p/shake-drivers/` (visited on 01/02/2012).

[65] Hyman, I., Boss, S., Wise, B., McKenzie, K. and Caggiano, J. Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology* 24.5 (2010), 597–607.

[66] International Telecommunication Union. World Telecommunication Development Report 2002. 2002. URL: `http://www.itu.int/ITU-D/ict/publications/wtdr_02/material/WTDR02-Sum_E.pdf` (visited on 01/12/2011).

[67] International Telecommunication Union. The world in 2011 – ICT facts and figures. 2011. URL: `http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf` (visited on 01/12/2011).

[68] Ishii, H. and Ullmer, B. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proc. CHI '97*, ACM (1997), 234–241.

[69] Jacob, R., Girouard, A., Hirshfield, L., Horn, M., Shaer, O., Solovey, E. and Zigelbaum, J. Reality-based interaction: a framework for post-WIMP interfaces. In *Proc. CHI '08*, ACM (2008), 201–210.

[70] Johnson, L. and Higgins, C. A navigation aid for the blind using tactile-visual sensory substitution. In *Proc. EMBS '06*, IEEE (2006), 6289–6292.

[71] Jones, M. and Marsden, G. Mobile Interaction Design. Wiley (2006).

[72] Jones, M. Journeying toward extravagant, expressive, place-based computing. *interactions* 18.1 (2011), 26–31.

[73] Jones, M., Buchanan, G., Cheng, T.-C. and Jain, P. Changing the pace of search: supporting "background" information seeking. *Journal of the American Society for Information Science and Technology* 57.6 (2006), 838–842.

[74] Jones, M., Buchanan, G., Harper, R. and Xech, P. Questions not answers: a novel mobile search technique. In *Proc. CHI '07*, ACM (2007), 155–158.

[75] Jones, M. and Jones, S. The music is the message. *Interactions* 13.4 (2006), 24–27.

[76] Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D. and Holmes, G. ONTRACK: dynamically adapting music playback to support navigation. *Personal and Ubiquitous Computing* 12.7 (2008), 513–525.

[77] Jylhä, A. and Erkut, C. A hand clap interface for sonic interaction with the computer. In *Proc. CHI '09: Extended Abstracts*, ACM (2009), 3175–3180.

[78] Kane, S., Wobbrock, J. and Smith, I. Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces. In *Proc. MovileHCI '08*, ACM (2008), 109–118.

[79] Karimi, H., Zimmerman, B., Ozcelik, A. and Roongpiboonsopit, D. SoNavNet: a framework for social navigation networks. In *Proc. LBSN '09*, ACM (2009), 81–87.

[80] Karvonen, K. and Parkkinen, J. A short note on the social consequences of wearables. In *Proc. SCI '01*, IIIS (2001).

[81] Kasemsuppakorn, P. and Karimi, H. Pedestrian network data collection through location-based social networks. In *Proc. CollaborateCom '09*, IEEE (2009), 1–9.

[82] Kindberg, T., Barton, J., Morgan, J., Becker, G., Caswell, D., Debaty, P., Gopal, G., Frid, M., Krishnan, V., Morris, H., Schettino, J., Serra, B. and Spasojevic, M. People, places, things: web presence for the real world. *Mobile Networks and Applications* 7.5 (2002), 365–376.

[83] Krüger, A., Aslan, I. and Zimmer, H. The effects of mobile pedestrian navigation systems on the concurrent acquisition of route and survey knowledge. In *Proc. MobileHCI '04*, LNCS 3160. Springer Berlin / Heidelberg (2004), 39–60.

[84] Kumar, A., Rajput, N., Chakraborty, D., Agarwal, S. and Nanavati, A. WWTW: the world wide telecom web. In *SIGCOMM '07 Workshop: Networked Systems for Developing Regions*, ACM (2007).

[85] Layar. Layar augmented reality browser. 2011. URL: `http://layar.com` (visited on 01/12/2011).

[86] Lee, J., Kim, Y. and Kim, G. Funneling and saltation effects for tactile interaction with virtual objects. In *Proc. CHI '12*, ACM (2012), 3141–3148.

[87] Li, K. Designing easily learnable eyes-free interaction. In *Proc. UIST '08 Doctoral Symposium*, Citeseer (2008).

[88] Li, K., Baudisch, P. and Hinckley, K. BlindSight: eyes-free access to mobile phones. In *Proc. CHI '08*, ACM (2008), 1389–1398.

[89] Lin, M.-W., Cheng, Y.-M. and Yu, W. Using tactons to provide navigation cues in pedestrian situations. In *Proc. NordiCHI '08*, ACM (2008), 507–510.

[90] Lindeman, R., Sibert, J., Mendez-Mendez, E., Patil, S. and Phifer, D. Effectiveness of directional vibrotactile cuing on a building-clearing task. In *Proc. CHI '05*, ACM (2005), 271–280.

[91] Löchtefeld, M., Schöning, J., Rohs, M. and Krüger, A. LittleProjectedPlanet: an augmented reality game for camera projector phones. In *MobileHCI '09 Workshop: MIRW*, 2009.

[92] Long, S., Kooper, R., Abowd, G. and Atkeson, C. Rapid prototyping of mobile context-aware applications: the Cyberguide case study. In *Proc. MobiCom '96*, ACM (1996), 97–107.

[93] Lucero, A., Boberg, M. and Uusitalo, S. Image Space: capturing, sharing and contextualizing personal pictures in a simple and playful way. In *Proc. ACE '09*, ACM (2009), 215–222.

[94] Luk, J., Pasquero, J., Little, S., MacLean, K., Levesque, V. and Hayward, V. A role for haptics in mobile interaction: initial design using a handheld tactile display prototype. In *Proc. CHI '06*, ACM (2006), 171–180.

[95] Lumsden, J. and Brewster, S. A paradigm shift: alternative interaction techniques for use with mobile & wearable devices. In *Proc. CASCON '03*, IBM Press (2003), 197–210.

[96] Ma, K.-T. and Peng, M.-C. Mobile phone and scent dispenser thereof. Patent Application US 2004/0235430 A1 (US). 25th Nov. 2004.

[97] MacKenzie, S. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7.1 (1992), 91–139.

[98] Magnusson, C., Rassmus-Gröhn, K. and Szymczak, D. The influence of angle size in navigation applications using pointing gestures. In *Proc. HAID '10*, LNCS 6306. Springer-Verlag (2010), 107–116.

[99] Mistry, P., Maes, P. and Chang, L. WUW – wear ur world: a wearable gestural interface. In *Proc. CHI '09: Extended Abstracts*, ACM (2009), 4111–4116.

[100] Munro, A., Höök, K. and Benyon, D. Personal and Social Navigation of Information Space. Springer-Verlag (1999).

[101] Murray-Smith, R., Williamson, J., Hughes, S. and Quaade, T. Stane: synthesized surfaces for tactile input. In *Proc. CHI '08*, ACM (2008), 1299–1302.

[102] Mustonen, T., Olkkonen, M. and Hakkinen, J. Examining mobile phone text legibility while walking. In *Proc. CHI '04: Extended Abstracts*, ACM (2004), 1243–1246.

[103] Nagel, S., Carl, C., Kringe, T., Märtin, R. and König, P. Beyond sensory substitution—learning the sixth sense. *Neural Engineering* 2.4 (2005), 13–26.

[104] Nicolai, T., Yoneki, E., Behrens, N. and Kenn, H. Exploring social context with the wireless rope. In *Proc. OTM '06 Workshops*, LNCS 4277. Springer Berlin / Heidelberg (2006), 874–883.

[105] Oakley, I. and Park, J.-S. Designing eyes-free interaction. In *Proc. HAID '07*, LNCS 4813. Springer-Verlag (2007), 121–132.

[106] Olofsson, S., Carlsson, V. and Sjölander, J. The friend locator: supporting visitors at large-scale events. *Personal and Ubiquitous Computing* 10.2–3 (2006), 84–89.

[107] Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proc. CHI '05*, ACM (2005), 919–928.

[108] Pascoe, J., Ryan, N. and Morse, D. Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction* 7.3 (2000), 417–437.

[109] Patel, N., Chittamuru, D., Jain, A., Dave, P. and Parikh, T. Avaaj Otalo: a field study of an interactive voice forum for small farmers in rural India. In *Proc. CHI '10*, ACM (2010), 733–742.

[110] Pearson, J. Investigating Lightweight Interaction for Active Reading in Digital Documents. PhD thesis. Swansea University, 2012.

[111] Petrie, H., Johnson, V., Furner, S. and Strothotte, T. Design lifecycles and wearable computers for users with disabilities. In *Proc. First Workshop on HCI for Mobile Devices*, Glasgow University (1998).

[112] Pielot, M., Henze, N. and Boll, S. Supporting map-based wayfinding with tactile cues. In *Proc. MobileHCI '09*, ACM (2009), 170–179.

[113] Pirhonen, A., Brewster, S. and Holguin, C. Gestural and audio metaphors as a means of control for mobile devices. In *Proc. CHI '02*, ACM (2002), 291–298.

[114] Poppinga, B., Pielot, M. and Boll, S. Tacticycle: a tactile display for supporting tourists on a bicycle trip. In *MobileHCI '09 Demos*, ACM (2009).

[115] Rapp, S., Michelitsch, G., Osen, M., Williams, J., Barbisch, M., Bohan, R., Valsan, Z. and Emele, M. Spotlight navigation: interaction with a handheld projection device. In *Pervasive '04 Videos*, 2004.

[116] Reitmayr, G. and Schmalstieg, D. Location based applications for mobile augmented reality. In *Proc. AUIC '03*, Australian Computer Society, Inc. (2003), 65–73.

[117] Rekimoto, J. Tilting operations for small screen interfaces. In *Proc. UIST '96*, ACM (1996), 167–168.

[118] Rekimoto, J., Miyaki, T. and Ishizawa, T. LifeTag: WiFi-based continuous location logging for life pattern analysis. In *Proc. LoCA '07*, Springer-Verlag (2007), 35–49.

[119] Rekimoto, J. and Nagao, K. The world through the computer: computer augmented interaction with real world environments. In *Proc. UIST '95*, ACM (1995), 29–36.

[120] Robinson, S., Eslambolchilar, P. and Jones, M. Point-to-GeoBlog: gestures and sensors to support user generated content creation. In *Proc. MobileHCI '08*, ACM (2008), 197–206.

[121] Rogers, Y. Moving on from Weiser's vision of calm computing: engaging ubicomp experiences. In *Proc. UbiComp '06*, LNCS 4206. Springer Berlin / Heidelberg (2006), 404–421.

[122] Rukzio, E. Physical Mobile Interactions: Mobile Devices as Pervasive Mediators for Interactions with the Real World. PhD thesis. LMU Munich, 2006.

[123] Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A. and Chin, J. An experimental comparison of physical mobile interaction techniques: touching, pointing and scanning. In *Proc. UbiComp '06*, LNCS 4206. Springer Berlin / Heidelberg (2006), 87–104.

[124] Schellenbach, M., Krüger, A., Lövdén, M. and Lindenberger, U. A laboratory evaluation framework for pedestrian navigation devices. In *Proc. Mobility '07*, ACM (2007), 495–502.

[125] Schilit, B., Adams, N. and Want, R. Context-aware computing applications. In *Proc. WMCSA '94*, IEEE (1994), 85–90.

[126] Schilit, W. A system architecture for context-aware mobile computing. PhD thesis. Columbia University, 1995.

[127] Seager, W. The Usability of Map-based Applications for Pedestrian Navigation. PhD thesis. Nottingham University, 2007.

[128] Seager, W. and Fraser, D. Comparing physical, automatic and manual map rotation for pedestrian navigation. In *Proc. CHI '07*, ACM (2007), 767–776.

[129] Sellen, A. and Harper, R. The Myth of the Paperless Office. MIT Press (2003).

[130] Sherwani, J. Speech Interfaces for Information Access by Low Literate Users. PhD thesis. Carnegie Mellon University, 2009.

[131] Shin, B.-S. and Lim, C.-S. Obstacle detection and avoidance system for visually impaired people. In *Proc. HAID '07*, LNCS 4813. Springer-Verlag (2007), 78–85.

[132] Simon, R. and Fröhlich, P. A mobile application framework for the geospatial web. In *Proc. WWW '07*, ACM (2007), 381–390.

[133] Simon, R., Fröhlich, P. and Anegg, H. Beyond location based – the spatially aware mobile phone. In *Proc. W2GIS '06*, LNCS 4295. Springer Berlin / Heidelberg (2006), 12–21.

[134] Simon, R., Frohlich, P. and Grechenig, T. GeoPointing: evaluating the performance of orientation-aware location-based interaction under real-world conditions. *Journal of Location Based Services* 2.1 (Mar. 2008), 24–40.

[135] Simon, R., Kunczier, H. and Anegg, H. Towards orientation-aware location based mobile services. In *Location Based Services and TeleCartography*, Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg (2007), 279–290.

[136] Simons, D. and Chabris, C. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28.9 (1999), 1059–1074.

[137] Sokoler, T., Nelson, L. and Pedersen, E. Low-resolution supplementary tactile cues for navigational assistance. In *Proc. MobileHCI '02*, LNCS 2411. Springer Berlin / Heidelberg (2002), 525–539.

[138] Soute, I. and Markopoulos, P. Head up games: the games of the future will look more like the games of the past. In *Proc. INTERACT '07*, LNCS 4663. Springer Berlin / Heidelberg (2007), 404–407.

[139] Starner, T. The enigmatic display. *Pervasive Computing* 2.1 (2003), 15–18.

[140] Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R. and Pentland, A. Augmented reality through wearable computing. *Presence: Teleoperators and Virtual Environments* 6.4 (1997), 386–398.

[141] Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S. and O'Modhrain, S. gpsTunes: controlling navigation via audio feedback. In *Proc. MobileHCI '05*, ACM (2005), 275–278.

[142] Strachan, S. and Murray-Smith, R. Nonvisual, distal tracking of mobile remote agents in geosocial interaction. In *Proc. LoCa '09*, LNCS 5561. Springer Berlin / Heidelberg (2009), 88–102.

[143] Strachan, S. and Murray-Smith, R. Bearing-based selection in mobile spatial interaction. *Personal and Ubiquitous Computing* 13.4 (2009), 265–280.

[144] Strachan, S., Murray-Smith, R. and O'Modhrain, S. BodySpace: inferring body pose for natural control of a music player. In *Proc. CHI '07: Extended Abstracts*, ACM (2007), 2001–2006.

[145] Strachan, S., Williamson, J. and Murray-Smith, R. Show me the way to Monte Carlo: density-based trajectory navigation. In *Proc. CHI '07*, ACM (2007), 1245–1248.

[146] Sugimoto, M., Ito, T., Nguyen, T. and Inagaki, S. GENTORO: a system for supporting children's storytelling using handheld projectors and a robot. In *Proc. IDC '09*, ACM (2009), 214–217.

[147] Thaler, R. and Sunstein, C. Nudge: Improving decisions about health, wealth, and happiness. Yale University Press (2008).

[148] The Economist. Not dead, just resting: how discredited technologies can be unexpectedly resurrected. 2008. URL: `http://www.economist.com/node/12381449` (visited on 01/12/2011).

[149] The World Bank. World development indicators: mobile cellular subscriptions. 2011. URL: `http://data.worldbank.org/indicator/IT.CEL.SETS.P2/countries/1W` (visited on 01/12/2011).

[150] Turkle, S. Alone Together: Why We Expect More from Technology and Less from Each Other. Basic Books (2011).

[151] United States Office of Compliance. Fast facts: the hazards of using mobile devices while walking. 2010. URL: `http://www.compliance.gov/wp-content/uploads/2010/06/Walking-Hazards-April-2010-Fast-Fact.pdf` (visited on 03/10/2010).

[152] Vainio, T. Exploring cues and rhythm for designing multimodal tools to support mobile users in wayfinding. In *Proc. CHI '09: Extended Abstracts*, ACM (2009), 3715–3720.

[153] Victor, B. A brief rant on the future of interaction design. 2011. URL: `http://worrydream.com/ABriefRantOnTheFutureOfInteractionDesign/` (visited on 08/11/2011).

[154] Want, R., Hopper, A., Falcão, V. and Gibbons, J. The active badge location system. *ACM Transactions on Information Systems* 10.1 (1992), 91–102.

[155] Want, R., Schilit, B., Adams, N., Gold, R., Petersen, K., Goldberg, D., Ellis, J. and Weiser, M. The ParcTab ubiquitous computing experiment. In *Mobile Computing*, The Kluwer International Series in Engineering and Computer Science 353. Springer US (1996), 45–101.

[156] Wasinger, R., Stahl, C. and Krüger, A. M3I in a pedestrian navigation & exploration system. In *Proc. MobileHCI '03*, LNCS 2795. Springer Berlin / Heidelberg (2003), 481–485.

[157] Weintraub, D. and Ensing, M. Human Factors Issues in Head-Up Display Design: The Book of HUD. Defense Technical Information Center (1992).

[158] Weiser, M. The computer for the 21st century. *Scientific American* 265.3 (1991), 94–104.

[159] Williamson, J., Strachan, S. and Murray-Smith, R. It's a long way to Monte Carlo: probabilistic display in GPS navigation. In *Proc. MobileHCI '06*, ACM (2006), 89–96.

[160] Williamson, J., Murray-Smith, R. and Hughes, S. Shoogle: excitatory multimodal interaction on mobile devices. In *Proc. CHI '07*, ACM (2007), 121–124.

[161] Williamson, J., Robinson, S., Stewart, C., Murray-Smith, R., Jones, M. and Brewster, S. Social gravity: a virtual elastic tether for casual, privacy-preserving pedestrian rendezvous. In *Proc. CHI '10*, ACM (2010), 1485–1494.

[162] Winograd, T. and Flores, F. Understanding Computers and Cognition: A New Foundation for Design. Addison Wesley (1987).

[163] Yi, B., Cao, X., Fjeld, M. and Zhao, S. Exploring user motivations for eyes-free interaction on mobile devices. In *Proc. CHI '12*, ACM (2012), 2789–2792.

[164] Young, M., Stanton, N., Walker, G., Jenkins, D. and Smart, W. Where do we go from here? An assessment of navigation performance using a compass versus a GPS unit. *Cognition, Technology & Work* 10.3 (2008), 231–236.